

```

DM of: US-09-886-400-3 to: SwissProt_40:* out_format : pfs
Date: Jun 11, 2002 10:41 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+np2 model -DEV=x1h
-O=cgn2+SWIPIOT+spool/US09886400/runat_11062002_150917-17471/app_quar
-DB=SwissProt_40 -CMIT=tastan -SUFFIX=rsp -GAPPB=12.000
-GAPEXT=0.000 -LOGPCL=0.000 -LOGPEXT=0.000
-QGAPEXT=0.100 -MINQFEXT=0.050 -XGAPOP=10.000 -IGAPEXT=0.500
-QGAPOP=4.000 -QGAPEXT=7.000 -YGAPOP=10.000 -IGAPEXT=0.500
-FGAPOP=6.000 -DEBLIX=7.000 -STDACT=1 -MATRIX=blosum62
-TRANS=human0 cdi -LIST=45 -DOCNAME=200 -THR SCORE=pct
-DELIM=6.000 -THR MIN=0 -ALGO=LOCAL -OUTFILE=pfs
-NORM=1.000 -UPDSTAT=500 -MINPEN=0 -MOTLEN=2000000000
-THR MAX=1.000 -THR MIN=0 -ALGO=LOCAL -OUTFILE=pfs
-NORM=1.000 -UPDSTAT=500 -MINPEN=0 -MOTLEN=2000000000

```

910 TTGAGGATATGGAGAGGACGAGGGAC.....GCAGACTAA 950
 ::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::
 316 Phe ..LeuTyrlsTyrProGluAsnTyrmethisArgLysAsn. 331
 951 TATGGTGTCTTACATAATGAGGGCAACTCGCCCTTTAGCGAGAGCA 1000
 ::::::::::::: ::::::::::::: ::::::::::::: :::::::
 331 uMetLeuSer.....LysLeuIleArgAsn. 340
 1001 GCGATGCCAAGGGATGGAGCCCTCCCTGAGGAGGGCTGGATGCCCT 1054
 ::::::::::::: ::::::::::::: ::::::::::::: :::::::
 341ProThrAlaArgIlePheValLeu 348
 1051 CGGGGGATATAAACGAT.....TGGAGGGCTGAAATGG 1086
 ::::::::::::: ::::::::::::: ::::::::::::: :::::::
 349 ArgAlaGlnCysAspAlaTyrTrpHisGlyValPheGly 362

seq_name: SwissProt_40.AMY1_DICTH
 seq_documentation_block:
 ID AMY1_DICTH STANDARD; PRT; 685 AA.
 AC P09961;
 DT 01-MAR-1989 (Rel. 1.0, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Alpha_amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase)
 GN AMYA.
 OS Dictyoglomus thermophilum.
 OX Bacteria; Dictyoglomus group; Dictyoglomus .
 NCBI_Taxid: 14;
 RN [1]
 RP SEQUENCE FROM N.A.
 STRAIN:H-6-12;
 RXN MEDLINE=83225059; PubMed=2453362;
 RAX Fukusumi S., Kamizono A., Horinouchi S., Beppu T.;
 RT "Cloning and nucleotide sequence of a heat-stable amylase
 from an anaerobic thermophile, Dictyoglomus thermophilum.";
 RL Eur. J. Biochem. 174:15-21 (1988)
 CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE: OL
 CC APPEARED AT THE BEGINNING OF INCUBATION, MAINTOSED AND GL
 CC DECREASE IN THE AMOUNTS OF MALTOOTROSE, MALTOTRIOSE AND GL
 CC PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE: THE OP
 CC TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 1 HO
 CC CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HO
 CC DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME.
 CC DEGREES CELSIUS.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucos
 CC linkages in Oligosaccharides and Polysaccharides.
 CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through
 CC between the Swiss Institute of Bioinformatics and the European
 CC Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content
 CC modified and this statement is not removed. Usage by an
 CC entities requires a license agreement (See <http://www.isb-sib.ch>).
 CC or send an email to license@isb-sib.ch.
 CC -----
 EMBL: X07896; CAA3075.1; -
 DR PIR; S00628; ALDAT.
 DR InterPro: IPR004300; Glyco_hydro_57.
 DR Pfam; PF03065; Glyco_hydro_57_1.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Multigen
 FT INIT_MET 0
 SQ SEQUENCE 685 AA; 81060 MW; C10941c8A508C404 CRC64;
 alignment_scores:
 Quality: 157.50 Length: 315
 Ratio: 52.063 Gaps: 15
 Percent Similarity: 22.540 Fingerprint: 315

alignment_block: US-09-886-400-3 x AMY1_DICTH .. .

Align seg 1/1 to: AMY1_DICTH from: 1 to: 685

22 GGCACCTCCGATGCCAAATCCCAAAGGGCAT 71
 ||||| :||:||:||:||:||:||:||:||:||:||:||:||:
 17 Glyasnpheasppheval.....TleGluArgIalaty 27

72 AGAGAGCCATACTACATCCGTCATCGAGAACCTGATAAAGAA.. GAAA 118
 ||||| :||:||:||:||:||:||:||:||:||:||:||:||:
 27 rGluMetSerTyrLysProLeuAsnLeuPheLysHsProLysP 44

119 TTCCGTTGGCTCACATAACGGCTATACCTTAAGGTCTCCGGAAAG 168
 :||:||:||:||:||:||:||:||:||:||:||:||:||:
 44 heProIleasnValHisPheSerCysLysLeuLeuLeuLeuPheLys 60

169CATATATAACCTGTTAAAGGGCCATGCCAGTGACCT 209
 ::||:||:||:||:||:||:||:||:||:||:||:||:
 61 AsnHisProGluTyrPheGluLysLeuSleMetAlaGluArgGly 77

210 GATAGAGATATCGAACACCACATACGACGATACTCCCCTCCCTCC 259
 ||||| :||:||:||:||:||:||:||:||:||:||:||:
 77 nileGluPheValSerGlyLysPheTyrGluProIleLeuProIleP 94

260 CGCTTAGCAGAGTAGAACGACAAGTCAGAGAGATAAGGAAGTAAGGAA 309
 ||||| :||:||:||:||:||:||:||:||:||:||:
 94 roAspLysPheValGlnGlnLysLeuAspLysTyrIleTyr 110

310 GAGCTCTTCGAGCTTCTCCAAAGGGATTCTGGCTAGGGCTGCCCTA 359
 :||| :||| :||| :||| :||| :||| :||| :|||:
 111 AspLysPheGlyLysIleProLysMetTrpIleAlaGluArgValTr 127

360 TGACCCGATATCCCTGCCATACCTGAGCAACAGCTATAGCTATCAT 409
 :||| :||| :||| :||| :||| :||| :||| :|||:
 127 PgluProHisLeuValLysTyrIleAlaGluAlaGlyIleGluTyralv 144

410 TCGCCACGGGAGCCGATGCTTCTCGCTCTAACCTGGCGATA 459
 :||| :||| :||| :||| :||| :||| :||| :|||:
 144 alvAlaLasp...AspAlaHisPheSerValGlyLeu..... 155

460 AAGCCAAATTAAACCGCTCATCCACACITATAAGGCCAAAGGAAAA 509
 :||| :||| :||| :||| :||| :||| :||| :|||:
 156ValPheProIleSerMetLysLeuArgTyrLeu..... 159

510 GCGCTTAGGTACATAGCTATCCCTGGCTAGGGGACTTAAGGAGG 559
 ||||| :||| :||| :||| :||| :||| :||| :|||:
 159 PhePheGlyTyr.....TyrLeuMetGluGluGlnGlyTyrIleU 174

560 CGATAAAGCTCGTTTGAAAGTAAAGTAACGCTAAAGGCACTAAAGAC 609
 ||| :||| :||| :||| :||| :||| :||| :|||:
 174 Ia.....ValPheProIleSerMetLysLeuArgTyrLeu..... 185

610 ATCGAAGCCGTACCCGTTGGGCCGTAACACGGCTGTAATGCTGG 659
 185 185

660 CATGGAAAGGCTTCCCTATGATCCTAAGAAGTGGCAGCTGATA. 708
 :||| :||| :||| :||| :||| :||| :||| :|||:
 186IleProPheAlaAspProGluGluThrIleThrTyrLeuA 199

709GAGGACAAAGGACAACATTCTATACGGACCCGAT 744
 ||| :||| :||| :||| :||| :||| :||| :|||:
 199 slysPheAlaSerGluAspIleSerLysIleAlaLeuLeuPheAspasp 215

745 ATGAGCTCATGGC.....TATAGGACATTCGAGGCTA 779
 ||| :||| :||| :||| :||| :||| :||| :|||:
 216 GlyGluLysPheGlyIleLeuTrpProAspPheAspThrIgthrVal.....Tyr 230

780 CAAATGAGTGTAGGGATATAGGGTT.....ATAGAGGAGCTCAACT 826
 :||| :||| :||| :||| :||| :||| :||| :|||:
 230 rGlu.....GlucGlyIleValAlaGlyPheValSerLysIleLysG 244

827 CGGAACTGTGCCCTCCCTCA.GAG 849
 :||| :||| :||| :||| :||| :||| :||| :|||:
 244 IuasnPhelLeuLeuValThrProIasnLeutYrrThrThrMetGlnArg 260

850 CTGAGGCAAGCTGGAGGGACTCTACTTACGGACTCTGAGTGG 894
 :||| :||| :||| :||| :||| :||| :||| :|||:
 261 VallysProLysGlyIarg.....TlePheProThrAlaSerTyr 274

seq_name: SwissProt_40_AMYA_PYRB

seq_documentation_block:

ID	AMYA_PYRB	STANDARD	PRT	655 AA.
AC	Q9V298;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Alpha-anhydase (EC 3.2.1.1).			
GN	AMYA OR PAB018.			
OS	Pyrococcus abyssi.			
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.			
NCBI_TaxID	23292;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAN-N-ORSAY;			
RA	"Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."			
RA	Heilig R.;			
RA	"Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution;"			
RT	Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.			
CC	-i- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucoosidic linkages in oligosaccharides and polysaccharides.			
CC	-i- PATHWAY: POLYSACCHARIDE DEGRADATION.			
CC	-i- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; AJ248283; CAB49100; 1. -			
DR	IntePro; IPR004300; Glyco_hydro_57.			
DR	Pfam; PF03065; Glyco_hydro_57.			
KW	Hydrolyase; Glycosidase; Carbohydrate metabolism; Complete proteome.			
SEQUENCE	655 AA; 7F6F920B1A00ECE CR064;			
alignment_scores:	Quality: 145.50	Length: 401		
	Ratio: 0.808	Gaps: 22		
	Percent Similarity: 44.888	Identity: 22.444		

255 CCTCCCGCTTAGAGAGTAGAAAGCACAAAGTCAGAGATAGGAAAGTTA 304
 ::|::|::|::|::|:
 92 alaProLysGluAspArgLeuGlyInletryLeuLysLutIpa 109
 ::|::|::|::|:
 305 AGGAAGAGCTTCTGGCTTCTCAAAAGGATCTGGTGCAGGCTC 354
 ::|::|::|::|:
 109 lalaSlys... GlyTyAspAlaLysGlyLeuLysLeuLysLutIpa 124
 ::|::|::|::|:
 355 GCGTATAACCGCATATAACCGCCATACAGGAAAGGTATGAGTA 404
 ::|::|::|::|:
 125 ValTrpGlnProGluLeuValLysThrLeuArgGluAlaGlyLeuGluTy 141
 ::|::|::|::|:
 405 TCTATTGCCGAGGGAGGGATGTTTCAGACTCATCTCAACTCTGG 454
 ::|::|::|::|:
 141 rValAlaAsp... AspTyRHisPheMetSerA 152
 ::|::|::|::|:
 455 CG...ATAAAGCCAATAAACCGCTCTATCCACCTATAAGGCCRAA 501
 ::|::|::|::|:
 152 IaGlyLeuSerLysAspGlnLeuPheTrpDtyTyrThrGluAspGly 168
 502 AGGAA...
 ::|::|::|::|:
 169 GlyGluValLeuThrValPheProLeaspGluLysLeuArgTyLeuI 185
 ::|::|::|::|:
 523ATCAGCTATCTCCTGGTCAGGG 547
 ::|::|::|::|:
 185 eProPheArgProValAspLysValLeuSerLeuHisSerUalAs 202
 ::|::|::|::|:
 548 AGCTTAGGAAGGGATAAGCTCTGGTTTGAAAGGTAAGGTAACGCTAAG 597
 ::|::|::|::|:
 202 erGluAspGluSerLysValAlaValPhenHis... 212
 ::|::|::|::|:
 598 GCAGTCAAAAGATCGAGGCCCTACCGCTGGGGCCSTGAAACAGGC 647
 ::|::|::|::|:
 213AspAspLysGluLysPheGlyLysLeuP... 221
 648 TGTATGTCGGCATCGGAAGGCTCTCTPATGATCCTAAAGAAAGTGG 697
 ::|::|::|::|:
 222Promet.....TheT 225
 ::|::|::|::|:
 698 CGAGCTGATAGGAGCAAGGC... 720
 ::|::|::|::|:
 225 yrGluTrpValTyGluLysGlyLysLeuValPhePheAspArgVal 241
 ::|::|::|::|:
 721AACATTCCTCTATACGGCACCGATATPAGAG... 750
 ::|::|::|::|:
 242 SerSerAspGluAlaLysAsnLeuMetLeuTySerGluLysLeuGlnTy 258
 ::|::|::|::|:
 751TTCATGGCTTATAGGACATGGTAGGCTACTCGGACTCTCCCTCAGA 848
 ::|::|::|::|:
 258 sphElysProLysGlyLeuValLeuProLeaSerItyPheGluM 275
 ::|::|::|::|:
 781AGATGAGTGTGAGGTTAGGGAA 798
 ::|::|::|::|:
 275 etserGluTrpSerLeuProAlaLysGlnAlaLysLeuPheValGlu... 290
 ::|::|::|::|:
 799 TTATAGGGTATAGCAGGTCAACTCGGAACTGTGCCTCTCCCTCAGA 848
 ::|::|::|::|:
 291 PheValGluLysLutIysGluLeuAsn... 299
 ::|::|::|::|:
 849 GCTGAAGCACAGTGAAGGGAGCTRACTTACGGACTTCGAGTTGGCAC 898
 ::|::|::|::|:
 300 ...MetPheGluArgTyArgValPheValArgGlyGlyIleTrp... 313
 ::|::|::|::|:
 899 CAGGAAAGCTTGAGGATATGGAGAAGGAGGAAC... 939
 ::|::|::|::|:
 314LysasnPhi... PhetyLysTyPheGluAlaAspTyNethis 327
 ::|::|::|::|:
 940 GCAAGACTTAATATGCTGTCCTACATAATGAGGGCGAACCTGCCTTT 989
 ::|::|::|::|:
 328 LysArgMetLeuMetLeuSer...ArgLeu 337

990 ACCCGAGAACAGGGATGGCAAGGGGATGGGCCCTCCCTGAGAGGGGC 1039
 ::|::|::|::|:
 337 UargAspAspProSerAlaArgGlyPhe... 346
 ::|::|::|::|:
 1040 TGGATGCGCTTCGGCGATATAACGAT....TGGAGGGGTGAAAT 1083
 ::|::|::|::|:
 347ValLeuArgAlaGinCysAsnAspAlaTyrTrpHisGlyValPhe 361
 ::|::|::|::|:
 1084 GGG 1086
 ::|::|::|::|:
 362 Gly 362
 ::|::|::|::|:
 seq_name: SwissProt_40:AMYA_METJA
 ::|::|::|::|:
 seq_documentation_block:
 ::|::|::|::|:
 ID AMYA_METJA STANDARD; PRT: 467 AA.
 ::|::|::|::|:
 AC 059000;
 ::|::|::|::|:
 DT 01-Nov-1997 (Rel. 35, Created)
 ::|::|::|::|:
 DT 01-Nov-1997 (Rel. 35, Last sequence update)
 ::|::|::|::|:
 DT 16-Oct-2001 (Rel. 40, Last annotation update)
 ::|::|::|::|:
 DE Putative alpha-amylase (EC 3.2.1.1).
 ::|::|::|::|:
 GN MJ1611.
 ::|::|::|::|:
 OS Methanococcus jannaschii.
 ::|::|::|::|:
 OC Archaea; Euryarchaeota; Methanococcaceae;
 ::|::|::|::|:
 NCBI_TaxID=2190;
 ::|::|::|::|:
 RN [1];
 ::|::|::|::|:
 SEQUENCE FROM N.A.
 ::|::|::|::|:
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 ::|::|::|::|:
 RX MEDLINE=95337999; PubMed=86880087;
 ::|::|::|::|:
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 ::|::|::|::|:
 RA Sutton G.J., Blake J.A., Clayton R.A., Gocayne J.D.,
 ::|::|::|::|:
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 ::|::|::|::|:
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 ::|::|::|::|:
 RA Scott J.L., Geoghegan N.S.M., Peterson J.D., Fuerstman J.L., Sadow P.W., Hanna M.C.,
 ::|::|::|::|:
 RA Utterback T.R., Kelley J.M., Hurst M.A., Borodovsky M., Cotton M.D., Roberts K.M.,
 ::|::|::|::|:
 RA Klener H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 ::|::|::|::|:
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 ::|::|::|::|:
 RT jannaschii.";
 ::|::|::|::|:
 RN Science 273:1058-1073(1996);
 ::|::|::|::|:
 CC -I-CRITICAL ACTIVITY: Endohydrolisis of 1,4-alpha-glucosidic
 ::|::|::|::|:
 CC linkages in oligosaccharides and polysaccharides.
 ::|::|::|::|:
 CC -I-SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 ::|::|::|::|:
 CC This SWISS-PROT entry is copyright. It is produced through collaboration
 ::|::|::|::|:
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 ::|::|::|::|:
 CC the European Bioinformatics Institute. There are no restrictions on its
 ::|::|::|::|:
 CC use by non-profit institutions in as long as its content is in no way
 ::|::|::|::|:
 CC modified and this statement is not removed. Usage by and for commercial
 ::|::|::|::|:
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 ::|::|::|::|:
 CC or send an email to license@isb-sib.ch).
 ::|::|::|::|:
 CC EMBL: U67601; AB990631.1; -.
 ::|::|::|::|:
 DR TIGR; MJ1611; -.
 ::|::|::|::|:
 DR InterPro; IPR004300; Glyco_hydro_57.
 ::|::|::|::|:
 DR Pfam; PF03065; Glyco_hydro_57; 1.
 ::|::|::|::|:
 KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;
 ::|::|::|::|:
 KW Complete proteome.
 ::|::|::|::|:
 SQ SEQUENCE 467 AA; 55558 MW; 40A6B1CDD4D967E CRC64;
 ::|::|::|::|:
 alignment_scores:
 ::|::|::|::|:
 Quality: 141.50
 ::|::|::|::|:
 Length: 377
 ::|::|::|::|:
 Gaps: 24
 ::|::|::|::|:
 Percent Similarity: 50.663
 ::|::|::|::|:
 Ratio: 0.741
 ::|::|::|::|:
 US-09-886-400-3 x AMYA_METJA
 ::|::|::|::|:
 alignement_block:
 ::|::|::|::|:
 Align seg 1/1 to: AMYA_METJA from: 1 to: 467
 ::|::|::|::|:
 40 GAATCCCCAAAAGGCGAAATCCCAAAGGTATAGAAGGCCATACATCCC 89

DR Pfam: PF03065; Glyco_hydro_57; 1;
 KW Transferase; Glycosyltransferase; Carbohydrate metabolism.
 SQ SEQUENCE: 653 AA; MW: 804651 F29219 CR64;
 alignment_scores:
 Quality: 121.50 Length: 386
 Ratio: 0.653 Gaps: 21
 Percent Identity: 21.244
 alignment_block:
 US-09-886-400-3 x MALQ_PYRK0 .
 Align seg 1/1 to: MALQ_PYRK0 from: 1 to: 653
 73 GAGAGGCATACGCCAGTCATGAGCACACTGATAAAGAAATTCC 122
 1||:||||| 1|| :||||| 1|| :||||| 1|| :||||| 1|| :||||| 1||
 28 GluArgSerTyrArgProPheMetGluThrLeu.....GluGlyLyrPr 42
 123 T.....TTGGCCTAACATAACGGCTATACCTTAAGTCCTC. 162
 42 CAsnMetLysValAlaValHistYSe-GlyProLeuGluTrpIle 59
 163CGAAAGATATTAGACCCTGTAAGGGCATCGGAGT 204
 59 ArgAsnLysProGluHisLeuAspLeuArgSerLeuValLysArg 75
 205 GACCTGATAGAGATAATCGGAACGAGCTAACCCACGAAATCTCCCT 254
 76 GlyGinLeuGluIleValAlaGlyPheTrpGluProValLeuAlaSe 92
 255 CCTCCCCGCTTAGCAGAGTAGAACGACAGTTCAGAGAGATAGGGAACTTA 304
 109 1aAsnAsnLeu...GlyTyGluIlaArgGlyValIleLeuGluGluArg 124
 92 rilePolySgluAspArgIleValGlnIleGluLeuGlyLeuGluPheA 109
 305 AGGAAGAGCTCTGAGCTTCTCAAGGGATTCTGCTGCCAGAGCTC 354
 109 1aAsnAsnLeu...GlyTyGluIlaArgGlyValIleLeuGluGluArg 124
 355 GCCTATGCCGATTAATCTGCCATACTGAGGACAACGGTTATGAGTA 404
 125 ValTrpGlnProGluLeuValLysSerLeuAlaAlaGlyLysAspTY 141
 405 TCTATTCGCGACGGGAGGGAGCTTCTCAGCTCATCTCAACTCGG 454
 141 rValLeuAsp.....AspTy-HisPheMetSerA 152
 455 CG...ATAAAGCCAATTAAACGGCTCATCCAAACCTTAAAGGCCAA 501
 152 LaGlyLeuSerLysAspGluLeuIleIhetPProTyTrpGluAspGly 168
 502 AGGAA.....AAGGCTTGTAGGTACATCG 527
 169 GlyGluValIleThrValPheProLeuAspGluIysLeuArg..... 182
 528 CTATCTCCTTGTCAGGGAGCTTAGGAGGCATAAGCTCGTTTT. 576
 183 .TyrLeuIleProPheArgProValAspLysThrLeuGluTyLeuHiss 199
 577GAAGGTAAAGTAAAGCCTAAAGCAGTC....AAAAGACATC 612
 199 erLeuAspAspGlyAspGluSerLysValIalayPheHisAspAspPly 215
 613 GAAGCCGTACCGTTGGTGGCCTGAACCGCTGFAATGCTGGCAT 662
 216 GlyLysPheGlyValIleTrpProGlyThrItyGluTrpValTyR..... 229
 663 CGGAGGGCTTCCTCTCTATGATCTAAGAACGTCGAGCTGATA.... 708
 230GlyLysGlyIysTrpIeuArgG 236
 709GAGGACAGGACAACTTCTCTATAC 735

236 IupHePheAspArgValSerSerAspGluArgIleAsnLeuMetLeuTyr 252
 736 GGZACCGATATAGAG.....TCATATGGCTATAGGGACAT 770
 :::: SerGluTyLeuGlnArgPheArgProArgGlyLeuValTyLeuProII 269
 771 TGCGAGCTAC...AGAATGAGGTTGAGGATTA..... 801
 ||||:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 269 eAlaSerTyrPheGluKetSerGluTrpSerIeuProAlaArgGinAlaL 286
 802TTAGAGGTATAGGGAGCTCATAGGACTTCGAACTGTCGATG 843
 286 YsLeuPheValGluPheValGluLeuLys..... 296
 844 TCAGAGCTGAAGCAGCACAGTGGAAAGGGCTCTACTAGGACTTCGAGTTG 893
 ::||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 297 LysIluAsnLysPheAspArgTyTrpArgValPheValArgGlyLysTr 313
 894 GGCACCAAGATAAGCTTGAGGATATGGAGAGGAGCAAGGGAAC..... 939
 | :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 313 P.....LysAsnPhe...PhePheLysTyrProGluSerAsnTyM 326
 | :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 940GCAAGACCTTAATATGCTGTCCTACAAATATGGGGGAACTCGGCC 984
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 326 eThiLysArgMetLeuMetValSerLysAlaValArgAsnAsn..... 340
 985 CTTITAGCCGAGAACAGCGATGCAAGGGGATGGAGCCCTCCTCCCTGAGAG 1034
 | :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 341PhePheLysIleLeuGluTrpIleLeuGluTrpIleLeuGluTrpIle 343
 | :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 1035 GAGGCTGGATGCCCTTCGGCGATATAACCAT.....TGAGGGGTG 1078
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 343 ArgGluPheIleLeuArgAlaGlnCysAsnAspAlaTyrrPhisGlyV 360
 1079 AAAATGGG 1086
 |||
 360 alpheGly 362
 seq_name: SwissProt_40:MALQ_THELI
 seq_documentation_block:
 ID MALQ_THELI STANDARD; PRT; 659 AA.
 AC 032462;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 4-alpha-glucuronotransferase (EC 2.4.1.25) (Amylomaltase).
 DE (Disproportionating enzyme) (D-enzyme).
 GN Thermococcus litoralis.
 OS Thermococcus litoralis.
 OC Archaea; Eurarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
 NCBI_TaxID=2265;
 RN [1]
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-19 AND 427-437, AND
 RC CHARACTERIZATION.
 RC STRAIN=DMS 54/3;
 RX MEDLINE=97454300; PubMed=9310375;
 RA Jeon, B.-S.; Taguchi, H.; Sakai, H.; Ohshima, T.; Wakagi, T.; Matsuzawa, H.;
 RT Thermococcus litoralis. Enzyme purification and characterization, and
 gene cloning, sequencing and expression in Escherichia coli.";
 RL Eur. J. Biochem. 248:171-178 (1997).
 CC -!- FUNCTION: Catalyzes the transglycosylation of
 CC maltooligosaccharides, yielding maltooligosaccharides of various
 CC lengths and glucose.
 CC -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucose
 CC to a new 4-position in an acceptor, which may be glucose or (1,4)-
 CC alpha-D-glucan.
 CC -!- ENZYME REGULATION: INHIBITED BY P-CHLOROMERCURIBENZOIC ACID,
 CC MONOODOACETIC ACID, MERCURY AND NICEL IONS.
 CC -!- MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isib-sib.ch/announce/> or send an email to license@isib-sib.ch).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).


```

OM of: US-09-886400-3 to: Issued_Patents_AA: * out_format : pfs
Date: Jun 11, 2002 10:34 PM
About: Results were produced by the GenCore software, version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+np.model
-DB=/com2_1/USPto_spool/US09886400/runat_11062002_150914-17298/app
-Q=0
-D=Issued_Patents_AA: * -QFMT=fastan -SUFFIX=rail -GATOP=12.000
-NINMATCH=0.100 -LOOPCL=0.000 -LOOPENT=0.000
-GAPEXT=4.000 -XGAPPOP=10.000 -XGAPEXT=0.500
-QGAPEXT=4.500 -QGAPPOP=6.000 -QGAPEXT=7.000 -QGAPPOP=10.000
-FGAPPOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-DELOPS=6.000 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -LIST=5
-NORMEXT=800 -NORMPOP=800 -MINLEN=0 -MAXLEN=2000000000000
-USRFILE=US09886400.GCGNI_1_31 -INCPU=6 -ICPU3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPY -WAIT -THREADS=1

Search information block:
Query: US-09-886400-3
Query length: 1095
Database: Issued_Patents_AA: *
Database sequences: 231628
Database length: 24425594
Search time (sec): 25.060000

```

33	hrLeuIleLysGluGlu... ProPheGlyLeuAsnLeuThrGlyTyrThr	48
151	TAAAGTCTCCGAAAGGATATTATAAGCCTCGTAAAGGGGATTCGC	200
151		151
49	LeuLysPheLeuProLysAspIle...LeuValLysGlyIleAl	64
201	GAGTGACTGTAGAGATAAOGAACGAGCTAACGCCGCAATTACTCC	255
201		201
64	aserAspLeuIleGluIleIleGlyThrSerItyrThr...AlaIleLeu	80
251	CCCTCCCTCCGTTAGGAGTAGAACAGTCAGAGATAGGGAA	300
251		251
80	roleLeuProIleSerArgValGluAlaGlnValArgAspArg...	95
301	CATAAGGAAGACTCTTCGAGCTTCTCAAAGGATTCTGGTCCAGA	350
301		301
96	ValLysGluIlePheGluValSerProLygIyPhenPheLeuProG	111
351	CCTCGCTATGCCGAGCCGATAATCCCTCCATGAGGACAACTGTTATG	400
351		351
112	ueuaIla...AspProIleProAlaLeuLysAspAsnGlyTyrG	128
401	AGTATCPATTGCCGAGGGAGGGATGCTTCTCAGCTCATCTCAAC	455
401		401
128	ItyrIlePheAlaAsp...GluIaMetIlePheSerAlaHistLeuAsn	14
451	TGCGCATAAACCAATTAAACGCCCTATCACACCTTATAAGGCCA	500
451		451
144	SerIleIleLysProIle...ProHisLeuIleIleSalaG1	155
501	AAGGAAAAGCCCTTAAGGTCATCTGCTATCCCTGGTCAGGAGC	555
501		501
159	NargG1IleSa9pheArgTyrIleSerItyrLeuLeu...LeuGluL	177
551	TGAGGAAGGCCATAAAAGCTGTTTGAAAGGTAAGCTAAAGGCCA	600
551		551
175	euArgLysAlaIleLeuIalPheIglGlyLysValIhrLeuIys...	199
601	GRCAAACATGGAGCCGTACCGCTTGGCTGCCGTGACACGGCTGT	655
601		601
191	ValLysAspIleGluAlaIvalProvaIrrpvalAlavalAsntrAlava	201
651	AATGCTCGATGCACTGGAAAGCCTCTATGAACTCTAAGAATGGGA	700
651		651
207	IMetLeu...IleGlyArgLeuProLeuMetAsnPolyLysValIalAs	222
701	GCTGGATAGGCCAAAGGACACAATCTCTATACGGCACCGATAATAGAG	755
701		701
223	erPheGluIspIys...AsnIleLeuLeuItyrIleItpIleGlu	231
751	TTGATTGGCTTATAGGGACATTCGAGCTACAGAATGAGTGTGAGGGATT	800
751		751
239	PhelGlyItyrArgAspIleAlaIgy...ArgMetrSerValGluIyle	255
801	ATAGAGGTTTATAGGACGCTTACGGCTACAGGACTGTCCTCCCTAGAGC	855
801		801
254	uleGluValIleAspGluLeuAsnSerGluIeucys...ProSerGluL	271
851	TGAAAGCAAGTGGAAAGGCCGCTACTTACGGACTCTGAGTGTGGGACCA	900
851		851
270	eWlShISeRgYArgGluIeuctyruSeuA9thrSerSeTrpAa...	281
901	GATAAGACCTGGCATTGGAGGACGAACTGCCCCTTTAGCCGAGAACAA	955
901		901
302	nMethIeu...TyrAsnMetArgIyGluLeuIalaphateuAgl1AsnS	311
286	AspIlysSeLeuIurgIietrPargI1aspGluG1AsnAlaArgIleus	301
951	TATGCTGTCCTACATAATATGAGGGCAGACTGCCCTTTAGCCGAGAACAA	100
951		951
1001	GCGATGCCAGGGATGGAGGCCCTCCCTGAGGGCTGCTGCTGCTTC	10
1001		1001
318	eraDAlaIargGlyTyr...ProIysProGluIargArgLeuAspAlaPhe	33

123 T.....TTGGGCTCAACATACGGCTATACCTTAAGTCCCTGCC 163
 43 oAsnMetLysValAlaLeHisThrSerGlyProLeuIleGluPheLeuG 60
 164 CG.....AAGGATATTAGACTCGTTAAAGGGGCATCCGAGT 204
 60 IlysPaspArgProGlyIleAspLeuIleArgSerLeuValLysArg 76
 205 GACCTGATAGATAATCGGAACGAGTACACGCCAACATACCCCCCT 254
 77 GlyInValGluIleValAlaLysPhenylGluProValLeuAlaSe 93
 255 CCTCCGCTTAGCAGTAGAGTCAAAGTCAGGAGATAAGGGAGTTA 304
 93 rilePro.....LysGluAspArgIleGluGinIleArgLeuMetL 107
 305 AGAA.....GAGCTCTTGAGCTTCTCAAAGGGAITCTGGCTGCCA 348
 107 ySGIutPralAllysSeriIgIpheAspAlaArgGlyValTrpLeuThr 123
 349 GAGGTCGCTTGTGAGGCGATAATCCCTGCCATACTGAAGACAACGGTTA 398
 124 GluArgValTrpGlnProGluIeValLysIleLeuLysGluSerGlyI 140
 399 TGAGTATCTATCGCGAGGGGATGCTTTCAGCTATCTCA 448
 107 ySGIutPralAllysSeriIgIpheAspAlaArgGlyValTrpLeuThr 123
 140 eAspTyryValIleValAsp.....AspTyryIlePhenM 151
 449 ACTGCCATAAACCCATTAAACCGCTCT.....CCACACCTTAAAG 495
 151 etserAlaGlyIeUserLysGluIleLeuIleGluIleGlyIleGlu 167
 496 GCCCAAAGGAA.....AAA 521
 168 AspIgIyGluValAlaValPheProIleAspIleGluIleAspIle 183
 522 CATGAGCTATCTCTGGTCAGGGAGCTTAGGAAGGGATAAGCTC. 570
 184TyrLeuIleProArgProValAspIleAspIleAspIleAsp 198
 571GTTTTGAGGTGTTGTTGTTAACGCTTAAGCCAGTC.AAA 606
 198 euHssSerIleIleAspIleAspIleAspIleAspIleAspIleAsp 214
 607 GACTCGAGGCCAACCGTTACCGTTGGTGGCCGTGAAACACGGCTGTTATGCT 656
 215 AspIgIyGluIlePheAspIleAspIleAspIleAspIleAspIle 230
 657 CGGATCGGAAGCCCTCCCTCTATGAATCTTAAGAAAGTGGCGACTGGA 706
 231GluIysGlyIleAsnLeuMet 235
 707 TA.....GAGCACAGGAACATCTT 729
 235 euArgGluIlePheAspArgIleSerAspIleIleAsnLeuMet 251
 730 CTATCGGCAACGATATAG.....TTCATGGCTATAG 764
 252 LeuIleIleAspIleAspIleAspIleAspIleAspIleAspIle 268
 765 GGACATTGGCTACAGAATGAGTTGAGGGATTAGAGGTATAG 814
 268 uProleAlaSerIle.....PheGluMet.... 276
 815 ACGAGCTCAACTCGAACCTGCTTCCC..... 843
 277SerGluIleProAlaLysGinAlaArgLeuPhe 289
 844TAGACTGANGCAAGTGA.AGGGA 869
 290 ValGluPheValAsnGluIleIleAspIlePheGluLysTyraR 306
 870 GCTCTACTTACGGACTTCAGTGGCACCATAGAGCTGAGGATAT 919

306 gValPheValArgGlyGlyIleIlePhe.....LysAsnPhe...PheT 319
 920 GAGAGAGGAGGAGGGAA.....GCAGACTATAGCTGTC 960
 319 YRLStyryProGluSerAsnTyrmethIleuMetValSer 335
 961 TACAATATGAGGGGAAACTGCCCTTTAGCCGAGAAACAGCAGTCAAG 1010
 336 LysLeuValArgAspAsn..... 341
 1011 GGATGGAGGCCCTGAGAGGCTGGAATGCCCTTCGGCGATAT 1060
 342ProGluIalArgLysTyrlleuArgAlaGinC 353
 1061 ATAACGAT.....TGGAGGGT 1077
 353 yAsnAspAlaTyrrIrrPhisGly 360

seq_name: /cgn2_6/ptodata/l/iaa/5A..COMB.pep:US-07-893-928A-1

seq_documentation_block:

; Sequence 1, Application US/07893928A
; Patent No. 5578479

; GENERAL INFORMATION:
; APPLICANT: LADERMAN, KENNETH
; ADDRESS: ANFINSSEN, CHRISTIAN
; TITLE OF INVENTION: a-AMYLASE FROM HYPERTHERMOPHILIC
; ARCHAEBACTERIUM
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0., version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/893, 928A
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95470/C-1197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-893-928A-1

alignment_scores:

Length: 391
 Percent Similarity: 46.803
 Identity: 22.762

alignment_block:

US-09-886-400-3 x US-07-893-928A-1 ..

Align seq 1/1 to: US-07-893-928A-1 from: 1 to: 650

73 GAGGGCATACATCCAGTCTGAGACACTGATAAAGAAATTCC 122
 29 GluIyscystyrrprPheLeuGluThrIeu.....GlutyltyrPr 43
 123 TTTGGGCTCARATAACGGCTATACCTTAAGGTCTCC 163
 1 :
 43 OASMetlysvalAlaHisthSerGlyProleuIleGluTrpLeu 60
 164 CGAAGGATATTAGACCTCGTTAAAGGGCATTCCGAGT 204
 60 IasnAspAsnArgProGlutylIleAspIleLeuArgSerLeuVallysArg 76
 1205 GACCTGATAGAGAATGGAACTGAGTAGAGAGATGGGAAGTTA 254
 255 CCTCCGCCTAGCAAGGACACAGTCAGAGAGATGGGAAGTTA 304
 93 rIIproLysGluAspArgIleValAlaLysGluIle 107
 305 AGGAA.....GAGCTCTTGAGCTTCTCCAAGGGATCTGGCTGCCA 348
 111 :
 107 ySgutrpAlaLysSerIleGlyPheAspAlaArgGlyValIleUthr 123
 349 GAGTCGCCTATACCCTATACCTGCCATACCTAAGGAAACGGTTA 398
 111 :
 124 GluArgValTrpGlnProGluLeuValLysThrLeuGluSerGlyIle 140
 399 TGAGATATCTATGCCAGGGAGATGCTTCTCACCTCTCTCA 448
 111 :
 140 EASTYrValIleAlaAsp.....AspTyrThsPheM 151
 449 ACTCGCGATAAACCAATAAACCGCTCAT.....CCACACCTTAAAG 495
 111 :
 151 etserAlaLysLeuSerLysGluLeuIleTrpProTyrThrGluIle 167
 496 GCCCARAGGGAAAGGCGCTTGGGTTAGGTA 521
 111 :
 168 AspGlyGluValAlaLysProIleAspGluIysLeuArg. 183
 111 :
 522 CATAGCTATCCTCTGGCTCAGGACTTGAAGCGATAAAAGCTC. 570
 111 :
 184TyrLeuLeuProAspLeuAspLysValIleGluItyrIle 198
 111 :
 571GTTTGAAGGTAAGGTAACGCTAACGGCAGTC.....AAA 606
 111 :
 198 euHisSerLeuIleAspGlyAspGluIysValAlaLysLeuArg.PheHisASP 214
 607 GACATCGAGGCCGTACCGGTTGGTGGCCGSGTAACACGGGTGTAATGCT 656
 111 :
 215 AspGlyGluLysPheGlyIleTrpProGlyThrTyrLysIlePheValTyR. 230
 657 CGGCATCGAAAGGGTCTCTATGAATCTAAGAATGCGGAGTGGGA 706
 231uProIleAlaSerIle.....GluIysGluItrpI 235
 707 TAGAGGACAAGGACAACTCTT 729
 111 :
 235 euarGluPhePhasPargIleSerAspGluLysIleAsnIleMet 251
 730 CTATACGGCACCGATATAGAG.....TTCACTGGCTPATAG 764
 111 :
 252 LeuThrIleThrLysLeuGluLysTrpProArgGlyLeuValItyrLe 268
 765 GGACATTCGAGGCTTACAGATGACTGTGAGGGATTTAGAGGTATAG 814
 111 :
 268 uProIleAlaSerIle.....PheGluMet.....PheGluMet... 276
 815 ACGAGCTCAACTGGAACACTGTGGCTTCCC..... 843
 111 :
 277SerGluItrpSerIleProAlaLysGlnAlaArgLeuPhe 289
 844TCAGAGCTGAAAGCACAGTGGGA.....AGGGAA 869

1057 SerPheAlaAspValProAlaLysAsp.....ArgValThrArg 1071
 75 GAAGGCATACATCCAGTCATGCAACACTGATAAA.....GAAGAAA 118
 1071 gGInGlyPheSerValleAspAspAlaSerIysPheAlaGlnGlnL 1088
 119 TTTCT.....TTGGCTCACACAAACGGCTATACCTTA 153
 ::||| euProGluLeuAsnSerPheGlySerIysIleAlaGlyAspLaser 1104
 154 AAGTCTCCGAAAGGATATAGACCTCGTAAAGGGCATGCCGAG 203
 1105 LysGluGlyProValAsnValLeuGlnVal.....GLYAlaLeuSerG1 1119
 204 TGACCTG...ATAGAGATAATC.....GGAACGGCTACGCCAAC 244
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 1119 YaspIresIleG1AspIeuG1LysAathSerAlaAsnlysASPL 1136
 245 TACTCCCTCTCCGCTTAGAGAGTAGAACACAAGTTCAGAGASAT 294
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 1136 ysLeuAsnMetIeuGlyIvaArgThrAlaLeuIleProArgGly 1152
 295 AGGAAGTTAAGGAAGGCTCTCCAGCTTCAGGCTTCCTCAAAGGGATCTGGCT 344
 ::|||:|||:|||:|||:|||:|||:|||:|||:
 1153 lysLys.....AspProSerTyrySerPh 1161
 345 GCGAGCTGCC.....TAGACCCGATAATCCCTGCCATACTGA 385
 ::|||:|||:|||:|||:|||:|||:|||:
 1161 eProGlnCySSerGlyPhethySgluAspProLeu..... 1172
 386 AGGACAACGGTTATGATATCTATTGCCGACGGGAGCGATGCTTTTC 435
 1172 1172
 436 TCAGCTCATCTCACTCGGCTATAAGCCATTAAACGGCTCATCCACA 485
 1173 ArgArgGlyMetArgPcmhrPhaHishi 1182
 486 CCTTATAAGGCCAAAGG...GAAAAGGCCCTT..... 516
 ::|||:|||:|||:|||:|||:|||:|||:
 1182 sLeuLeuGluLeuGlyArgLeuGluAsnPheLeuGluArgIleP 1199
 517 AGCTACATCGCTATCTCTGGCTCTAGGGAGCTT... 552
 ::|||:|||:|||:|||:|||:|||:
 1199 roAlaValGlyArgAsnValGlnIleTerPheGlySerGluLystrhala 1215
 553 ...AGGAAGGCCATAAGCTGTCTTGAAAGTAGGTAAACGTAAGC 599
 ::|||:|||:|||:|||:|||:|||:
 1216 ArgArgAsnAlaAlaGlnValValhe.....LeuArgAl 1227
 600 AGTCAAAAGACATGAAAGCGTACCCGTTGGGCCGTTGAAACGGCTG 649
 ::|||:|||:|||:|||:|||:
 1227 alIleSerHistIleProGlyLeuThrThrPheSerGlyAlaArgAlaL 1244
 650 TAATGCTGGCATGGAAAGCTCTCTTATGAACTTAAGAAGTGGCC 659
 ::|||:|||:|||:|||:|||:
 1244 euLeuGlyLeuAspGluLeuGluArgAlaAlaasnSerLysVal 1260
 700 AGTGGATAGAGCAAGGAAACATTCCTCATACGGC..... 738
 ::|||:|||:|||:|||:|||:
 1261 Ser...ValGlnSerSerArgIleTerPheLeuProGluG1 1276
 739 ACCGATA..... 1290
 ::|||:|||:
 1276 nSeAspAlaThrProGluGluLeAlaLysGluHeGluGly.....
 767 ACATGGAGGCTACAGAATGAGTGTGAGGGATATTAGGGTTAGAC 816
 1291 1293
 ::|||:|||:
 817 GAGCTCAACTCGGAACTGTGCTTCCTCAAGCTGAAGCACAGTGGAG 866
 ::|||:|||:
 1294 LysLeuIysSerArgLeu.....AlaGlnAr 1302

867 GGAGCTCTACTTACGGACTTCGAGTTGGCACCGATAAGACCTTGAGGA 916
 ::|||:|||:|||:|||:|||:
 1302 gLeuIhrLysIleArgValAspGlu.....IleGluThrLysValArgV 1317
 917 TATGGAGAGGGAGCAA.....GGAAACGCAAGA.....CTTAATATG 954
 ::|||:|||:|||:|||:
 1317 alThrValGlnAspGluAspGlySerProArgValValProValArgIeu 1333
 955 CTGCTCTACATATAGGGCGAA 978
 ::|||:|||:|||:|||:
 1334 ValAlaSerSerMetGlnGlyGlu 1341
 seq_name: /cgn2_6/podata/1/1aa/5A_COMB.pep:US-08-418-893D-24
 seq_documentation_block:
 / Sequence 24 , Application US/08418893D
 / Patent No. 5559220
 / GENERAL INFORMATION:
 / APPLICANT: ROESSLER, PAUL G
 / APPLICANT: OHLROGE, JOHN B
 / TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
 / TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
 / NUMBER OF SEQUENCES: 25
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
 / STREET: 1617 Cole Blvd.
 / CITY: Golden
 / STATE: CO
 / COUNTRY: USA
 / ZIP: 80401-3393
 COMPUTER READABLE FORM:
 / COMPUTER: IBM PC compatible
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC DOS/MS-DOS
 / SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/418,893D
 / FILING DATE: April 7, 1995
 / CLASSIFICATION: 800
 PRIORITY APPLICATION DATA:
 / APPLICATION NUMBER: US 08/104,938
 / FILING DATE: September 14, 1993
 / CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 / NAME: O'CONNOR, EDNA
 / REGISTRATION NUMBER: 29,252
 REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
 TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 303-231-1000
 / TELEFAX: 303-231-1098
 / TELEX:
 / INFORMATION FOR SEQ ID NO: 24:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 2089 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / HYPOTHETICAL: NO
 / ANTI SENSE: NO
 / FRAGMENT TYPE: N-terminal
 US-08-418-893D-24

alignment_scores:
 Quality: 88.50
 Ratio: 0.497
 Percent Similarity: 49.721
 Percent Identity: 19.553

alignment_block:
 US-09-886-400-3 x US-08-418-893D-24 ..

```

817 GAGCTCAACTCGGAACTGCGCTTCCTGAGGTGAAAGCACAGTGAG 866
:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
1294 LysLysLysSerArgLeu.....AlaGlnAr1302

867 GGCGCTACTTAGGACTTCGAGTTGGCACCAAGATAGAGCTTGAGGA 916
|:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
1302 gLeuthrLysLeuArgValAspGlu.....IleGluThrLysValArgV 1317

917 TATGGAGAGGGACCAA...GGGAAGCAGAAGA.....CTTAAATAAG 954
:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
1317 althRValGlnAspGluAspGlySerProArgValValProValArgLeu 1333

955 CTGCTCTACATATAGGGCAGA 978
:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
1334 ValAlaSerSerMetGlnGlyGlu 1341

seq_name: /cgn2-6/ptodata/1/iaa/_6B_COMB.pep;us-09-355-166-1
seq_documentation_block
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Generon International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1

alignment_scores:
Quality: 87.50
Ratio: 0.57
Percent Similarity: 47.147
Length: 333
Gaps: 19
Percent Identity: 22.823
Identity: 73

alignment_block:
US-09-355-166-1 .. .
Align seg 1/1 to: US-09-355-166-1 from: 1 to: 657
84 CATCCCGTATCGAGAACACTGTATTAAGAAAGAAAATTCCCTTGGGCTCA 133
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
61 HisGlyGluLysArgSerThrAspProArgTrpSer.Pro..... 73
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
134 ACATACGGCTATACCTTAAGTCTCCCGAAG.....GATATT 174
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
74 ....AspGlyArgThrLeuAlaLeuSerAspArgGluGlyAlaPala 88
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
175 ATAGACCTC.....GTAAAGGGCATGGCAGTGACCTGAT 212
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
89 AlaGlnLeuTyrosLysSer-ThrGluGlyGluAlaArgLysLeuTh 105
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
213 AGAGATA...ATCGGAAACGACCTAACAG.....CACG 241
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
105 RAspLysProTyroGlyValSer-LysProLeuTrpSerProAspGlyGlu 122
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
242 CAATGGGAAGTAAAGGA.....GAGCTCTTCGAGGTTC 291
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
122 erileuLeuValThrIleSerLeuGlyGluGlyGluSerIleAspPapArg 138
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
292 GATAGGGAAAGTAAAGGA.....GAGCTCTTCGAGGTTC 326
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
139 GluLysThrGluLysAspSerIysGluProValGlyValGlyLeuSer 155
|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
327 TCCAAG.....GAGCTCTTCGAGGTTCGAGCTGAGCTATGACCCGATAA 370

```

155 RTyRLysArgAspGlyLysGlyLeuThrArgGlyAlaTyRAlaGlnLeu 172
 371 TCCCTGCCATACTGAG
 :: :::::|||:
 172 alieuValSerVallysSerGlyGluMetLysGluLeuThrSerHisLys 188
 388 ...GAAACCGTTTGTAGTAGTATCATTGCGGAGGGAGGCATGCUTTT 434
 189 AlaAspGlyIysApproAlaPheserProAspDlyLysTrpLeuValPh 205
 435 CTCAGTCATCAACTCGCGATAAGGCC
 |||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 205 eSerAlaAsnIleuThrGlutHAspAspLaseLysSroHsAspValT 222
 466ATTAACCGCTCTATCCACAC... 486
 :::::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 222 yRlMetSerIeuGluSerGlyAspLeuLysGlnValThrProHisArg 238
 487CTTATAAAAGCCCCAAAGGAAAAGCGCTTTAGTACATAGCTA 530
 :::::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 239 GlySerSerGlySerSerHeserProAspDlyArgTyrlleAile 255
 531 TCTCTCTGGCTCTAGGGAGCTTAGGAAGCCA 562
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 255 uLeuGlyAsnGluLysGluLysSerAlaThrLeuSerLysAlaIrrPL 272
 563 TAAAGCTGTTTTGAGGTAAAGC...CTAAAGGAGTCAAAGAC 609
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:
 272 eutyrAspIleGluGlnGlyArgLeuThrCysLeuThrGluMetLeuAsp 288
 610 ATCGAACCCGTAACCCCTTGGTGCCTTATGAAATCCCTAAAGAAAPGTCGGTAGTGCTGG 659
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:
 289 ValHisLeuIaAspAlaLeuIleGlyAspSerLeuIleGlyI 303
 660 CATGGGAGGGCTTCCCTTATGAAATCCCTAAAGAAAPGTCGGTAGTGCTGG 709
 |:|||:|||:|||:|||:|||:|||:|||:
 303 YaaGluGlnIargProlle....
 710 AGGACAAGGACAACATCTCTATAAGGGACCGATAATAGAGTCATGGC 759
 ::|||:|||:|||:|||:|||:|||:|||:|||:
 312 SAPsPsrGlySerGlyPheVtIleGlyThrasPglnglySerThrGly 328
 760 TATGGGACATGCACTGAGCTCAACTGGAACATGTGAGGATTTAGGGT 809
 |||:|||:|||:|||:|||:|||:|||:|||:
 329 lle.....TyrTyrlleSerIleGluGlyLeuValTyPr 340
 810 TATAGAGGAGTCAACTGGAACATGTGAGGATTTAGGGT 859
 |||:|||:|||:|||:|||:|||:
 340 oile ..ArgLeuGluLysGlu..... 346
 850 GTGGAAGGGAGCTTACTTACGGCTTCGAGTTGGCACCAGATAAG 906
 |||:|||:|||:|||:|||:|||:
 347TyrIleAsnSerPheserLeuUserProAspGlu 357

seq_name: /cgn2_6/prtdata/1/1aa/5B_COMBO.pep;US-09-141-135-2

seq_documentation_block:

Sequence 2, Application US/09141135
 Patent No. 5981729
 GENERAL INFORMATION:
 APPLICANT: LEE, Yong Hun
 TITLE OF INVENTION: Transcription Factor Gene Induced by Water
 FILE REFERENCE: 1942/31
 CURRENT APPLICATION NUMBER: US/09/141,135
 CURRENT FILING DATE: 1998-08-27
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: WordPerfect 6.1/Windows
 SEQ ID NO 2
 LENGTH: 235
 TYPE: PRT
 ORGANISM: Arabidopsis thaliana

Patent No. 5648240
GENERAL INFORMATION:
 APPLICANT: Hook, Magnus
 APPLICANT: Jonsson, Klas
 APPLICANT: Patti, Joseph M.
 APPLICANT: Gurusiddappa, Sivasankarappa
 TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: United States of America
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/248,021A
 FILING DATE: 24 MAY 1995
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: TAMK:155
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-1000
 TELEFAX: (512) 474-7577
 INFORMATION FOR SEQ ID NO: 2:
 .SEQUENCE CHARACTERISTICS:
 LENGTH: 689 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-248-021A-2

Alignment_start	Alignment_end	Sequence	Qualitative_scores	Qualitative_scores	Length:	Length:
1	271	Length: 271 GAGACACTGATTAAGGAGAAATTCCCTTGGCTAACATAACGGCTA 146	86.50	86.50	271	271
1	11	Gaps: 11 Percent Identity: 18.819	0.765	0.765	11	11
1	11		41.697	41.697	11	11
1	165				165	165
1	297	Align seg 1/1 to: US-08-248-021A-2 from: 1 to: 689 rthrhrAsnIleGlnSerAsnLeuIapheserAsnLysproThrhrA 297	97	97	297	297
1	280	Length: 280 GluIalaylsValAsnGlnValProTyrsSerIleAsnLeuAsnGlyl 280	166	166	280	280
1	330	Length: 330 GlyValSerGluArgAspLeuLysHisAlaLysPaspValIleAsnSer 330	314	314	330	330
1	347	Length: 347 TyrPhelysAsnGlyGlyLysArgValIleHisAlaLysPaspIleIva 347	208	208	347	347
1	322	Length: 322 CTGATAGAGAATCGGAACGGCT 322	233	233	322	322
1	347	Length: 347 TAGCAAGTAGAACCAA 347	347	347	347	347

Page 1

```

DM of: US-09-886-400-3 to: A_Genesed_032802:*
out_format : pfs
Date: Jun 11, 2002 10:33 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+nlp.model -DEV=xlh
-Q=/cgn2/_/USP01-spool1/US09886400/runat_1106002_150913_17285/app
-DB=A_Genesed_032802 -QEXT=fastan -SUFFIX=rag -GAOP=11_000
-MINMATCH=4_000 -LOOPCL=0_000 -LOOPXT=0_000
-QGAPEXT=0_000 -XGAPEXT=0_000 -XGAPDXT=0_000
-QGAPOP=4_500 -XGAPEXT=0_050 -YGAPEXT=0_000
-FGAPOP=6_000 -FGAPEXT=7_000 -YGAPEXT=0_500
-DELN=6_000 -DELEXPT=7_000 -START=1 -MATRIX=tBlosum62
-TRANS=human40_cdi -LIST=45 -DOCALLIGN=200 -TR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN15 -MODE=LOCAL -OUTFILE=pfs
-NORMEXT=HESSIZE=500 -MINLEN=2000000000
-USERUS09886400_@CGN1_1_0 -NCPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLUPXY -WAIT -THRADS=1

Search information block:
Query: US-09-886-400-3
Query length: 195
Database: A_Genesed_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 53.920000

```

/SIDS1/gcgdata/genessq/geneseqp-embl/AA2001.DAT : AAMB0176 + 91.00 159.82 0.2944
 /SIDS1/gcgdata/genessq/geneseqp-embl/AA2001.DAT : AAB62559 + 91.00 159.82 0.2944
 /SIDS1/gcgdata/genessq/geneseqp-embl/AA2001.DAT : AAG82313 + 91.00 157.99 0.3186
 /SIDS1/gcgdata/genessq/geneseqp-embl/AA2001.DAT : AAG62551 - 91.00 152.72 0.3995
 /SIDS1/gcgdata/genessq/geneseqp-embl/AA2001.DAT : AAM15156 + 91.00 148.33 0.4833

 seq_name: /SIDS1/gcgdata/geneseqp-embl/AA1997.DAT:AAW34643

 seq_documentation_block:
 ID AAW34643 standard; Protein: 364 AA.
 XX
 AC AAW34643;
 XX DT 27-MAR-1998 (first entry)
 XX DE Thermostable alpha-galactosidase AEDIII2RA-alpha-gal-18GC.
 XX
 KW Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;
 KW food processing; alpha glycosidase hydrolysis; raffinose;
 KW starchose; verbascose; bean; flatulence; AEDIII2RA-alpha-gal-18GC.
 XX OS Thermococcus alcaliphilus strain AEDIII2RA.
 XX
 Key Location/Qualifiers
 FH FT Msc-difference 329
 FT /note= "encoded by CTT"
 XX PN WO9732974-A1.
 XX PD 12-SEP-1997.
 XX PF 05-FEB-1997; 97WO-US01452.
 XX PR 08-MAR-1996; 96US-9613220.
 XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
 XX PI Murphy D, Reid J, Rudolph MJ;
 XX DR WPI; 1997-470541/43.
 XX N-PSDB; AAT93753.
 XX
 PT Nucleic acid encoding alpha-glycosidase from Thermococcus
 PT alcaliphilus - used in food processing to hydrolyse
 PT alpha-glycosides, e.g. raffinose
 XX PS
 CC Claim 1: Fig 1; 32pp; English.
 XX
 CC This protein comprises AEDIII2RA-alpha-gal-18GC, a claimed
 CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
 CC AEDIII2RA, a bacterium that shows optimum growth at 85 deg C and
 CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
 CC AAT93753) encoding the alpha-galactosidase; (2) a vector containing
 CC the polynucleotide or homologous or complementary sequences; (2)
 CC host cells containing the vector; (3) a process for producing the
 CC alpha galactosidase in transformed or transfected host cells; an
 CC enzyme showing at least 70% identity to alpha galactosidase and
 CC comprising at least 30 amino acid residues of its sequence; and (4)
 CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
 CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
 CC in sugar beet processing, raffinose inhibits crystallisation of
 CC sucrose, and as a digestive aid to hydrolyse raffinose, stachyose
 CC and verbascose in beans and other gassy foods
 XX Sequence 364 AA;

 alignment_scores:
 Quality: 1873.00 Length: 364
 Ratio: 5.146 Gaps: 0
 Percent_Similarity: 100.000 Percent_Identity: 99.725
 alignment_block:

US-09-886-400-3 x AAW34643 ..

Align seg 1/1 to: AAW34643 from: 1 to: 364

1 TTGAGGGCTCGCTTCAGGGAAACCTCCAGTATGCCGAATCCTCAA 50
 1 LeuArgAlaLeuValPheHisGlyAsnLeuGlnTrpAlaGluLeuProly 17

51 GAGCAAATCCAAAGCTCATAGAGGAGGATACTACCCAGTCATCGAGA 100
 17 SerGluLeuProlySarLysAlaGluLysAlaGluLeuProly 17

101 CAGTATTAAGAAATCCTTGGCTCAATAAGGCTATACC 150
 34 hrLeuLysGluGluLysProhepLeuAsnThrGlyTyrThr 50

151 TTAAGITCTCCGAAAGGATATTAGACCTCGTAAAGGGGATTCGC 200
 51 LeuIysPheLeuProlySarAspLysAspLeuAspLysGlyGlyIleAl 67

201 GAGTGACTGTAGAGATAATGGAACGCTTACCGCAGCAATACCTC 250
 67 aserAspLeuLysGluLysGlyThrSerTyrThrHisAlaLeuP 84

251 CCCCTCTCCGCTTACCGAGACTAGAACAGTCAGAGAGATGGAA 300
 84 ronLeuLeuProLeuSerAspLysAspLeuAspLysGlyGlyIleAl 100

301 GTAAAGGAGAACCTTCAGAGTATCCAAAGGGATTCTGGCAGCCAGA 350
 101 ValIysGluGluLeuPheGluLeuSerProlyGlyPheTrpLeuProG 117

351 GCTGCCATGACCGATAATCCGCCATACTGAGGAAACGGTTARG 400
 117 uleAlaLysRapProLysLeuLysAspAsnGlyTyrc 134

401 AGPATCTATTCCCGACGGGNGCCATGCTTCTCAGCTCACTCAAC 450
 134 luteLysPheAlaAspGlyGluAlaMetLeuPheserAlaHisLeuAsn 150

451 TCCGGATAAACGCCATAAACGCTTAAAGGCCCA 500
 151 SerAlaLysProLysProLeuTyProHsLeuLysAlaG 167

501 AAGGGAAAAGGGCTTGGTACATCAGCTATCCTGGCTCTCACGGAGC 550
 167 nargGluLysArgPheArgTyRisertyLeuLeuGluGlu 184

551 TTAGGAAGGCGATAAACGCTTGGTAAAGGTAACGCTTAAAGGCCA 600
 184 euargLysAlaLysLeuValPheGluGlyLysValThLeuIysAla 200

601 GTCAGAACATGACGGCTAACGGTACCGCTTGGCTCTCACGGCTG 650
 201 ValIysAspLysAlaAlaAlaProValTrpValAlaAsnThrAlaVa 217

651 ATGCTCGCATGAGGCTTCCCTCTTGAATGAAAGTGCAGACGGCTG 700
 217 MetLysLysLysLysLysArgLeuProLeuMetAspProLysValAla 234

701 GCGGATAGGGCAAGGACACATCTCTCATGGCACGGATATAGAG 750
 234 erTriPheGluLysAspAspAspAspAspAspAspAspAspAspLysGlu 250

751 TCTATGGCTATAGGGCATTCAGGCTACAGAATGAGTGTGAGGGATT 800
 251 PheIegGlyTArgAspPheLeuLysGlyTArgMetSerValGluGlyLe 267

801 ATTAGAGGTTATAGACCSAGCTAACCTGGAACTGCGCCCTCTAGAGC 850
 267 uLeuGluValLeuAspGluLeuAspGluLeuAspGluLeuAspGluLeu 284

851 TGAAGCAGTGGAAAGGGCTTACTTACGGACTTCGAGTTGGGCACCA 900

284 eulyShisSerGlyArgGluLeuTyLeuArgThrSerSerTPalPro 300

901 GATAAGGCTGAGGATATGGAGAGGAGAAGGAGACTAA 950
 301 AspLySSerLeuArgLysLeuPheGluAspGluGlyAslAlaArgLeuAs 317

951 TAGCCTGCTCTACAATAATGGGGGGAACTCTGCCCTTCTAGCCGAGACA 1000
 317 ntLeuLeuSerTyrAspMetArgGlyGluLeuAlaPheLeuAlaGluAns 334

1001 CGGATGCAAGGGATGGGAGGCCCTCCCTGAGAGGAGGTGGATGCCRTC 1050
 334 erAspAlaArgDiyTrpGluProLeuProLysGargLeuAspAlaPhe 350

1051 CGGGCATATAAACATTGGGGTGAANNTGGGACCT 1092
 351 ArgGluLeuTyrAspTLPargGlyGluAnsGlyGluPro 364

seq_name: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT: AAB96088

seq_documentation_block:

ID AAB96088 standard; Protein: 655 AA.

XX

AC AAB96088;

XX

DT 29-OCT-2001 (first entry)

XX

DE Putative P. abyssi alpha-amylase.

XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX

OS Pyrococcus abyssi.

XX

PN FR2792651-A1.

XX

PD 27-OCT-2000.

XX

PF 21-APR-1999;

XX

AX 99FR-0005034.

XX

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

PI Quereiou J, Weissenbach J, Saurin W, Heiling R;

XX

DR WPI; 2001-126236/14.

XX

PA (CNRS) CNRS CENT NAT RECH SCI.

PA (IFREMER) IFREMER INST FR RECH EXPL MER.

XX

XX

PS Claim 7: Pages 711-712; 1657BP; French.

XX

CC The present invention relates to the genomic sequence of Pyrococcus abyssi (see AA86431 and AAB1123-7) and P. abyssi proteins. P. abyssi is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present sequence is one such P. abyssi protein.

CC The proteins of the present invention have various potential industrial uses, since the proteins are stable at very high temperatures, some up to CC 110 degrees centigrade.

CC Note: This patent is in the same patent family as WO2000065062, which CC contains additional sequences as shown in AAB9132-AAB99143,

CC AAB75003-AAB7590 and AAB6436.

XX

Sequence 655 AA;

alignment_scores: Quality: 145.50 Length: 401

Ratio: 0.808 Gaps: 22

Percent Similarity: 44.888 Percent Identity: 22.444

SQ Sequence 649 AA;

alignment_scores:
 Quality: 143.50 Length: 399
 Percent Similarity: 0.776 Gaps: 24
 Percent Identity: 46.366

alignment_block:
 05-09-886-4-00-3 × AAR47504 ..

Align seg 1/1 to: AAR47504 from: 1 to: 649

73 GAGAGGCCATACATCCAGTCAGTCAAGACACTGATAAAGGAGATTCC 122
 29 GluIysCstYrTrpProLeuGluThrLeu.....GluGluTyrP 43

123 T.....TTGGCTCAACATAACGGGTATACTTAAGTCCCTC 163
 43 OASnMetLysvalAlaLeHisThrSerGlyProLeuIleGluTrpLeu. 59

164 CGAGGATT.....ATAGACCTCGTTAAAGGGGCATTCGCG 201
 60 ..GinAspIleArgProGluTyrlleAspLeuArgSerLeuIlys 75

202 ATGACCTGTATAGAGATAATGGAAAGGAGTACAGCCACCAACTCCC 251
 76 ArgGlyInvaGluIleValAlaGlyHypTrpIleProLeuI 92

252 CCTCTCCGCTTAGCAGAGTAGAACACAAGTCAAGAGATAGGGAAG 301
 .92 aSerIlePro.....LysGluIaspArgIleGluGinIleArgLeuM 106

302 TTAGGAA.....GAGGCTCTGAGTTTCCTCAAAGGGATTGGCTG 345
 106 etyGluIutrpAlaIlysSerIleGlyPheAspAlaArgGlyValTrpLeu

346 CCAAGGCPGCCATGACCGATAATCCCTCCATAGGGAAACGG 395
 123 ThreGluArgValTrpGlnProLeuValIlysThrLeuIlysGluSerG 139

396 TTATGAGTATCTATTCGCCGACGGGGCATGGCTTTCTCAGCTCATC 445
 139 YIleAspIryValIleValAsp...AspTyFHissPheMetSerAlaGluL 155

446 TCAACTCG..... 453
 155 euSerLysGluIgluLeuIrrTrpProTyrrTrpIleAspGlyGlu 171

454GGGATAAAAGCCAATTAACCGCTCTATCCACACCTTAAAGGC 497
 172 ValIleAlaValIpheProIleSp..... 179

498 CCAAGGAAAGCAGCTTAGTACATCAGTATTCCTGCTGCTAGGG 547
 180GluIlys.....LeuArgTrpIleProheArgP 190

548 AGCTTAGGAGGCCATAAGCNC.....GTTTGAGGTAAAGGT 588
 190 roValAspIysValIpheGluIleUHissLeuIleAspGlySpGlu 206

589 ACGGTAAAGCAGTC.....AAAGACATCGAAGCCGTACCGTTGGGT 632
 207 SerIysValAlaIleAlaIpheHisAspAspGlyGluIysPheGlyIleTrpPR 223

633 GGCGTGAAACACGGCTGAAATGCTGGCATCGAAAGCTTCCTATGA 682
 223 OGlyIrrTrpGluIrrPvItyr..... 230

683 ATCTTAAGAAAGTCGAGCTGCGATA..... 708

231GluIysGlyIrrTrpIleArgGluPhePheAspArgIleSer 243

709 ...GAGGACAAGGACAACATTCCTCTATAGGCCACCGATAATAGAG..... 750
 244 SerAspIluIysIleasnIeumIeutyrrhrIluIysLeuGluIlysT 260

751TCATGGCTPATAGGACATTCGAGCTACAGATGAGTG 790
 260 rIysProArgGlyIleValItyrLeuProIeaIaserIyr..... 273

791 TICAGGGATTATAGGTATAGCTGACTCGAAACTGCGCTT 840
 274PheGluMet.....SerIluTrpSerIeu 281

841 CCC..... 857
 282 ProAlaIysGlnAlaArgLeuIpheValGluPheValAsnGluIeuIysva 298

858 CAGTGG.....AGGACCTCTACTATGGATTCGAGTGG 895
 298 IlysGlyIlePheGluIlysTrpArgValIpheValArgGlyGluIleTrp. 314

896 CACCAGATAAGGCTGAGGATATGGAGACAGGAGAAC..... 939
 315LysAsnPie...PheTyryLysTrpProGluSerAsnTyryNet 327

940 ..GCAAGACCTAAATATGCTGTCCTACAATATGAGGGCGAACATCGCCCT 986
 328 HisLysArgMetLeuMetValSerIysLeuValIargAsnHn..... 341

987 TTAGGCCGAGAACAGGGATGCAAGGGGATGCCCTCCCTGAGAGGA 1036
 342 345
 1037 GCTGGATGCTTCCGGCCTATATAACCAT.....TGGAGGGT 1077
 345 rIlysTrpIleuIargAlaGlnCysAsnAspAlaTyryTrpIhsGly 360

seq_name: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT: AAB96091

seq_documentation_block:
 ID AAB96091 standard; Protein: 1362 AA.
 XX
 XX AAB96091;
 AC
 XX DT 29-OCT-2001 (first entry)
 XX DE Putative P. abyssi amylolytulanase.
 XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
 OS XX Pyrococcus abyssi.
 XX FR2792651-A1.
 XX PD 27-OCT-2000.
 XX PF 21-APR-1999; 99FR-0005034.
 XX PR 21-APR-1999; 99FR-0005034.
 XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O; Querellou J, Weissenbach J, Saurin W, Heiling R;
 XX DR WPI: 26001-126236/14.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFREMER INST FR RECH EXPL MER.
 XX PT New nucleotide sequences isolated from Pyrococcus abyssi encode.
 PT proteins useful in industry.
 XX PI
 XX Claim 7: Pages 715-719; 1657pp; French.
 PS XX
 CC The present invention relates to the genomic sequence of Pyrococcus

abyssi (see AAF86431 and AAH41223-7) and P. abyssi is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present sequence is one such P. abyssi protein. The proteins of the present invention have various potential industrial uses, since the proteins are stable at very high temperatures, some up to 110 degrees centigrade.

Note: This patent is in the same patent family as WO2000065662, which contains additional sequences as shown in AAB99132-AA899143, AAH75903-AAH75920 and AAG66436.

Sequence 1362 AA;

alignment_scores:

Quality: 124.00 Length: 381

Ratio: 0.756 Gaps: 16

Percent Similarity: 43.045 Percent Identity: 21.522

alignment_block:

US-09-886-400-3 x AAB96091 ..

Align seg 1/1 to: AAB96091 from: 1 to: 1362

94 ATCAGAGACTGATTAAGAAGAAATTCCCTTTGGCTAACATAACGGG 143

:::|||||:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

236 ValGluIleValLeuLysGlnMet...TrpLeuAsnHisThr.. 250

144 CTATACCCTAAAGTCCCTCCGARRGGATTATAAGCTCGTTAAAGGGG 193

:::|||||:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

251PheGluGluHisGluLysIleAsnLeuLeuGly. 262

194 GCATGCCGAGTGACTGTAGATAGAGATAATCGGAACGAGCTAACGCCAGCA 243

:::|||||:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

263AsnGlyAsnValGluValThrValAlaP-totyThrHisPro 276

244 ATATCCCOCTCCCG.....CTTAGCAGAGTAGAGNCACA 281

|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

277 IleGlyProIleLeuAsnAspPheGlyTrpIleGlyAspPheAspIleGly 293

282 AGTCAGAGAGATAGGGGAGTAACTGGAAACGAGCTAACGCCAGCA 322

:::|||||:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

293 nvalLysLysAlaAsnGluLeuIrrlyGlyTrpIleGlyAlaGlyLysV 310

323 TTTCCTCAAAGGGATTGCGCAGAACGCTCATGACCGATAATC 372

:::|||||:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

310 althrProIleGlyGlyTrpAlaAlaGluSerAlaLeuAsnAspIleThr 326

373 CCTGCCATACTGAGGAAACGSGTATGAGTCTATTCGCCTACGGG 420

|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

327 LeuGluIleAlaGluAsnGlyTrpIleGlyAlaGlyLysV 343

421GAGGGCTATGCTTCTCACTGCCTCATCAACTCGCGATAAACGG 463

:::|||||:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

343 uvalLeuGluLysLeuGlyValProLyserIleGlySerIleGly 360

464 CA. 465

360 rotPrValAlaGlnPheGlyAspLysIleIlePheProArgAsn 376

465 465

377 HisAspLeuSerAspArgValGlyPheArgTyraAlaGlyMetAsnInTy 393

466ATTAACCGCTCATCACACCTTATAAGGCCRAAGGGAA 508

393 rAspAlaValLysAsnPheValGluLeuLeuLysIleGlnLyGlnA 410

509 AGCGCTTTAGGTACATCAGCTATCCTGGCTCAGGACGCTTGAAG 558

410 snrYAspGlySerLeuValTyraValIleThrLeuAspGlyGluAsnPro 426

559 GCGATAAACCTCGTTTGAAGSTAAGCTAACGCTAAAGGAGTC...AA 605

|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

New 4-alpha -glucanotransferase which has been heat-treated - used at

427 TrpGluHisTyrProProheAspGlyLysLeuPheLeuGluGluLeuTyrAr 443

PT

606 AGACATGAAAGCGTACCGTTGGTGGCTGACAGGCTGA... 651

PR

443 gGinLeuGluGluLeuGlnLysLysGlyLeuIleArgThrValThrPro 460

XX

652ATGCTCCG... .ATCGGAGGCTTCCTCTTANG 681

PF

460 er-GluTyryIleLysMetPheGlyAspLysAlaAsnLysLeuThr-ProLys 476

XX

682 ATCCCTAGAAGTGGCAGCTGGATAGAGCACAGACATCTCT 731

PR

477 MetMetDysA9LeuAspPheThrIleGluAsnAlaLeuLe 493

XX

732 ATACGGCACCGATAATAGAGTCATGGCTATAGGACACATGGAGCTACA 781

PR

493 uysAlaIlysIleLeu.....GlyGluL 501

XX

782 GAATGAGTGTGAGGGATTATAAGGTTATAGAGCTATAAGGAGCTCAACTGGAA 831

PR

501 eutYAspMetValGlyValIhrGlu..... 509

PT

832 CTTGGCTTCCTCTCAAGCTGAAGCAGGAGCTACTTACG 881

PR

510GluMetGlnTrpPr 514

XX

882 GATTCGAGTGTGGCACCAGATAAGAGCTTGGAGATATGGAGAGGGAGC 931

PR

514 GluIusSerItpIle... ASPGlyThrLeuSerItpPr... 527

PT

527 leGlyGluProGlnGluAsnIleAlaTrpPrTrpLeuIleAlaArg 543

PT

982 GGCCTTATGGGAGAACGCAAGCTGAGGAGCCCTCCC 1028

PR

544 LysAlaLeuPheGluAsnLysAsnValAspItpAsnLysAlaIty 560

PT

1029 TGAGGGAGGCTGGATGCCCTCGGGATATAACGATG 1071

PR

560 rGluYr.....LeuPheArgAlaGluGlySerAspItp 571

PT

seq_name: /SIDS1/gcqdata/geneseq/geneseqp-embl/AA1998.DAT:AAW54870

PT

seq_documentation: /SIDS1/gcqdata/geneseq/geneseqp-embl/AAW54870 standard; Protein: 653 AA.

ID

AAW54870; AC

XX

01-SEP-1998 (first entry)

XX

Super heat resistant 4-alpha-glucoanotransferase.

XX

Super heat-resistant 4-alpha-glucoanotransferase; heat-treatment;

KW

alpha-1,4-giucan; alpha-1,4-giucoside bond.

KW

Pyrococcus sp.

OS

JP10150986-A.

XX

PN

PD

09-JUN-1998.

XX

21-NOV-1996; 96JP-0311117.

PP

21-NOV-1996; 96JP-0311117.

XX

(BEBB-) BE ABLE KK.

PA

(NAGA-) NAGASE SEIKAGAKU KOGYO KK.

XX

WPI; 1998-379989/33.

DR

N-PSDB; AAV7026.

XX

PT high temperatures to transfer at least one glucose unit
 XX Claim 4; Page 10-13; 18pp; Japanese.
 XX

The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of 6.0-8.0, with an optimum temperature at pH 7.5 of 100 deg. C. It has at least 90% activity after heat treatment at 100 deg. C for 30 minutes at pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glucan to alpha-1,4-glucan by an alpha-1,4-glucoside bond.

CC SO Sequence 653 AA;

alignment_scores:
 Quality: 121.50 Length: 386
 Ratio: 0.63 Gaps: 21
 Percent Similarity: 48.187 Percent Identity: 21.244

alignment_block:
 US-09-886-400-3 x AAW54870 ..

Align seg 1/1 to: AAW54870 from: 1 to: 653

73 GAGAGGGATACTCCCGATCATCGAGACACTGATAAAGAAAGAATTCC 122
 |||||:|||||:||| :||| :|||:|||:|||:|||:|||:|||:|||:|||:|||
 28 GluArgSerTyPArgProPheketGlutHrleu.....GluGlutYPr 42

123 T.....TGTGGCTCAACATAACGGCTATAACCTTAAGTCTC. 162
 | :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 42 OasmMetLysValAlaValHistYsSerGlyProLeuLeuGlutPilea 59

163CCGAAGGATATTAGAGCTCGTAGAACGGGCATCCCGAGT 204
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 59 rgAAPASLysPfOgluHisLeuAspLeuLeuArgSerLeuValYsArg 75

205 GACCTGATAGAGATAATGGGACGAGCTACACGCCAGCAATACTCCCCCT 254
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 76 GlyGlnLeuGluIleValAlaGlyPhenYrgIleProValLeuAlaSe 92

255 CCTCCCGCTTAGAGAGTAGAAAGCAAAAGTTCAGIGAGATAAGGGAGTTA 304
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 92 rleProLysGluAspArgIleValGlyLeuLysGluIlePha 109

305 AGGAAGAGCTCTCGAGCTTCTCAAAGGGATTCTGGCTGCCAGAGTC 354
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 109 laArgAsnLeu...GlyTyrGluAlaArgGlyValTrpLeuThrIleArg 124

355 GCCTATGCCCATATCCCCTGCCAAACTGAGGGCAACCGTTATGAGTA 404
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 125 ValTrpGlnProGluLeuValIleSerLeuArgAlaAlaGlyIleAspY 141

405 TGTATTCCCGACGGGAGGGATGCTTCTCACGCTCATCTCACACTCG 454
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 141 rValIleValAsp.....AspYrrHisPheMetSerA 152

455 CG...ATAAGCCATTAAACCGCTCTATCCACACCTTATAAACGCCAA 501
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 152 IaglyLeuSerLysAspGluLeuIlePheTrpProTyrThrGluaspGly 168

502 AGGAA.....AAGCCCTTAAAGGTACATCG 527
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 169 GlyGluValIleThrValPheProIleAspGluIleLeuGlyLeuHis 182

528 CTATCTCTTGGCTTCAGGGAGCTAGGAAGGGATAAGCTCGTTTT. 576
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 183 .TyrLeuIleProPheArgProValAspLysThrLeuGlyLeuHis 199

577GAAGTAAAGTAACGCTAACGGCAGTC.....AAAGACATC 612
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 199 erLeuAspAspGlyAspGluSerLysValAlaPheHisAspAspGly 215

613 GARGCCGACCGTATGGCTGGAAACACGGCTGTAATGCTGGCAT 662
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

216 GluLysPheGlyValTrpProGlyThrYrGluTrpValYr 229
 663 CGAAGGCTTCTCCCTPATGATCTAAAGTGGCGAGCTGGATA..... 708
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 230GluLysGlyTriPheLeuArg 236
 CC 709GAGGACAAAGGACAACATTCTCTATAC 735
 CC 236 luPheAspArgValSerSerAspGluIargLeasnLeuMetLeuYr 252
 CC SO 736 GGACCCGATATAGAG 770
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 253 SerGluTyRLeuGlnPheArgProArgGlyLeuValYrLeuProIle 269
 771 TGCAAGGCTAC ..AGAATGAGTGTGAGGGATA 801
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 269 eIaLsErTyPheGluMetSerGluTrpSerLeuProAlaArgGlnAlaL 286
 802TTAGAGTTATAGAGGACTAAACTGGAAACTGTGCTCTCC 843
 286 ySleuPheValIciLpheValGluLeuLeuLys 296
 844 TAGAGCTGAAAGCACAGTGAAAGGAGCCTACTTACGACTGAGTTG 893
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 297 LysGluAsnLysPheAspArgTyPArgValPheValArgGlyIleTr 313
 894 GCACCAAGATAAGAGTTGAGATGGAGAGAGACGAGAGGGGAC 939
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 313 PLysAsnPhe...PhePhelYstYsProGluSerAsnTyRm 326
 940GCAAGACTTAATAATGTTGTCCTACATAATGAGGGGCAACTCGCC 984
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 326 ethIlysArgMetLeuMetValSerLysAlaValArgAsnAsn 340
 985 CTTTATGCCGAAACGCGATGCAAGGGATGGGCCCTCCCTGAGG 1034
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 341PheGluIleLeuGlyLeuLeuGlyLeuLeuLeuLeuLeuLeu 343
 1035 GAGGCTGGATGCTTCGGGGGATATAAACATTGAGGGGG 1078
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 343 argGluPheIleLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyV 360
 1079 AAAATGGG 1086
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 360 alPheGly 362

seq_name: /SIDS/geo/data/geoquery/geneseq/emb1/AA2001.DAT:AA982528
 seq_documentation:
 block: ID AA8B228 standard; Protein: 360 AA.
 XX AAG8B228;
 AC XX
 AC XX
 XX DT 03-SEP-2001 (first entry)
 XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
 XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
 XX DE S. epidermidis SR1 strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX Staphylococcus epidermidis.
 OS XX
 PN XX
 WO200134809-A2.
 XX PD 17-MAY-2001.
 XX
 XX PR 09-NOV-2000; 2000WO-US30782.
 XX PR 09-NOV-1999; 99US-0164258.
 XX PA (GLAXO) GLAXO GROUP LTD.
 XX PI Kimmerly WJ;

PT reagent contg. protein, NADH and L-aspartic and alpha-keto-glutaric
acid for determination of glutamine oxalo-transaminase activity
XX
PS Claim 8; Page 13-15; 17pp; Japanese.
XX This sequence represents a protein having heat resistant malate
dehydrogenase (h-rMAD) activity. The protein has a residual activity
after storage at 40 deg C for 10 days of at least 60%, pref. 70%
and esp. 90%. A reagent containing the h-rMAD protein, NADH and L-
aspartic and alpha-keto-glutaric acid may be used for the
determination of glutamine oxalo-transaminase (GOT) activity. The h-
rMAD protein may be produced by transforming E. coli with the DNA
encoding this protein and isolating the protein from the culture medium.
CC
XX Sequence 329 AA;

alignment_scores:
 Quality: 104.00 Length: 154
 Percent Similarity: 51.299 Gaps: 7
 Percent Identity: 25.974
 alignment_block:
 US-09-886-400-3 x AAR94013 ...
 Align seg 1/1 to: AAR94013 from: 1 to: 329
 28 CTCAGATGCCGAATCCCAAAGAGCGAAATCCCAAAG..... 66
 130 MetThrIrrhrValPhyLysIuSerglycHeProllysAsnArgValI 146
 67GTCATAGAAGGCCATACTCCCGTCAATCGAGACAC 103
 146 eglyGlnserGlyValleuAspthrAlaArg.....PheArgThrP 160
 104 TGTATANGAGNATCCTCTTGGGTCACTAACGGCTATACCTTA 153
 160 heValAlaGluIu.....LeuAsnIleSer..... 168
 154 AAGTCCCTCCGCCAGGGATATTAGACCTCGTTAACGGGCATCCGAG 203
 169VallysASPvalThrGlyPheValLeuGlyGlyHISGLyAs 182
 204 TGAGCTGATAGATAATGGAAACGACTCACGCCAGCAATA.....C-247
 182 PAsPMetValProLeuValArgTyrSerTyrAlaGlyGlyIleProLeuG 199
 248 TCCCCCTCTCCGGCTTAGCAGAGTAGAAGCAACAAGTCAGAGATAGG 297
 199 luryLeuIleProLeuAspArgLeuAspAlaLeuValGluArgThrArg 215
 298 GAAGTTAGGAAGAGCCTTCGAGCTTCTCCAAAGGA....TTCTG 341
 216 LysGlyGlyGlyIleLeuValLeuGlyAsnDlySerAlaTyrTyr 232
 342 GCTGCCAGAGCTGCCATTGACCGATAATCCCTGCCATACTGAAAGACA 391
 232 rAlaProAlaAlaSerLeuValGluMetValAlaIleLeuIlyAspG 249
 392 AC..... 399
 249 InArgArgIleLeuProAlaIleAlaTyrLeuGluGlyGlutYglYtyr 265
 400 GAGTATCTTTC 411
 266 GluGlyIleTyr 269

XX 17-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana Protein fragment SEQ ID NO: 16524.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 20000EP-0301439.
 XX PR 25-FEB-1999; 990US-0121825.
 PR 05-MAR-1999; 990US-0123180.
 PR 09-MAR-1999; 990US-0123548.
 PR 23-MAR-1999; 990US-0125788.
 PR 25-MAR-1999; 990US-0126264.
 PR 29-MAR-1999; 990US-0126785.
 PR 01-APR-1999; 990US-0127462.
 PR 06-APR-1999; 990US-0128234.
 PR 08-APR-1999; 990US-0128714.
 PR 16-APR-1999; 990US-0129845.
 PR 19-APR-1999; 990US-0130077.
 PR 21-APR-1999; 990US-0130449.
 PR 01-APR-1999; 990US-0130510.
 PR 23-APR-1999; 990US-0130891.
 PR 28-APR-1999; 990US-0131449.
 PR 30-APR-1999; 990US-0132048.
 PR 30-APR-1999; 990US-0132408.
 PR 04-MAY-1999; 990US-0132484.
 PR 05-MAY-1999; 990US-0132485.
 PR 06-MAY-1999; 990US-0132486.
 PR 06-MAY-1999; 990US-0132487.
 PR 07-MAY-1999; 990US-0132863.
 PR 11-MAY-1999; 990US-0134556.
 PR 14-MAY-1999; 990US-0134219.
 PR 14-MAY-1999; 990US-0134221.
 PR 18-MAY-1999; 990US-0134310.
 PR 18-MAY-1999; 990US-0134468.
 PR 19-MAY-1999; 990US-0134441.
 PR 20-MAY-1999; 990US-0135124.
 PR 21-MAY-1999; 990US-0135333.
 PR 24-MAY-1999; 990US-0135679.
 PR 25-MAY-1999; 990US-0136021.
 PR 27-MAY-1999; 990US-0136321.
 PR 01-JUN-1999; 990US-0136782.
 PR 03-JUN-1999; 990US-0137222.
 PR 04-JUN-1999; 990US-0137538.
 PR 07-JUN-1999; 990US-0137532.
 PR 08-JUN-1999; 990US-0138094.
 PR 10-JUN-1999; 990US-0138540.
 PR 14-JUN-1999; 990US-0139119.
 PR 16-JUN-1999; 990US-0139421.
 PR 17-JUN-1999; 990US-0139433.
 PR 18-JUN-1999; 990US-0139456.
 PR 18-JUN-1999; 990US-0139457.
 PR 18-JUN-1999; 990US-0139458.
 PR 18-JUN-1999; 990US-0139459.
 PR 18-JUN-1999; 990US-0139460.
 PR 18-JUN-1999; 990US-0139461.
 PR 18-JUN-1999; 990US-0139462.
 PR 18-JUN-1999; 990US-0139463.

seq_name: /SIDS1/gogdata/geneseq/geneseqp-embl/AA2000.DAT:AAG16038

seq_documentation_block:

ID_AAG16038 standard; Protein: 571 AA.

AC AAG16038;

220 ATCGAACGAGCTACACCAGGATACTCCCCCTCCGGCTAGCAG 269
 147 LeuIrgGinArgLeuThrLysThrMetLeuProLeu.....LeuSerG1 161
 270 AGTGAAGCCAAGTCAGAGAGAT.....294
 161 uLeuglyGlyIleLeuProArgGluThrLeuLeuTrpLeuLeuL 178
 295ACGAAAGTTAGGAA 309
 178 euArgSerGlyCysAlaTyraAsnSerArtLeuHisAlaValGlnAla 194
 310 GAGTCCTTCGAGCTT..TCCTCCAAGGGATTCTGCTGCCAGAGCTGC 356
 195 GluLeuIleLeuAlaSerGlyLysaspMetLeuProSerGlnG1 211
 357 CTATGACCCGATATCCCTGCCATACTG.....384
 211 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 228
 385 ..AGAGCACACGGTATGATGATCTGCTATCTGCGCGGGAGCCGAGCTT 432
 228 heLySAspAsnGlyIleIleLeuLeuGluAspSerIleSerLeuLeu 244
 433 TTCTAGCTCATCACTGGccATAAGCCATTAACCCCTATCC 482
 245ThrValIleNysGlyThrGly.....251
 483 ACACCTTATAAGGCCAAAGGGAAAAGGCTTTAGGTACATGC... 528
 252Lys Tyr Arg Ser Gln Trp Arg Trp Asp Leu Val Ser Asp P 265
 529TATCTCCCTGGT.....CTC 543
 265 heLeuProProSerIysGlyGlutLeuAlaLeuAspGluValIeu 281
 ,544 AGGGAGCTTAGGAAGCGATAAACCTCTTTT.....GA 578
 282 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrNetGluAs 298
 579 AGTAAAGGTAAACGCTAAAGGCAGTCANGACATCGAAGCCGTAACC 624
 298 pgIyIysIle.....ValIysGlyLeuAlaGlyValProAspPL 311
 625GTGGGTGCCGAAACCGCTGTAATGCTGGCATCGGAGG 669
 311 ysGlyProValLeuLeuAlaGlyTyrHisMetLeuMetGlyLeuGluLeu 327
 670 CTTCCCTTATGAAUCCCTAAAGAAATGGCGAGCTGATGGCAAGGA 719
 328 GlyProMetSerGlu.....AlaPhenIleLysGluIys... 338
 720 CAACATTCCTCTA.....732
 339 .AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrsAspAsnA 355
 733TACGCCACCGATAAGASTCATGGCTFAT 762
 355 sPProAlaLysAlaPheAspIysGlyAspTrpIleLysValPhgIy... 370
 763 AGGGACATTGAGGCTACAGAAATGAGTGTGAGGGATTAGAGGTTAT 812
 371AlaIysProValThrAlaIthrAsnLeuPhelysLeuE 383
 813 AGACGAGCTCACTCGAACTGTGCTTCCTCTCA.....GAGCTGA 853
 383 uAspSerLeuSerHisValLeuLeuPheProGlyIalaArgGluAlaL 400
 854 AGCACAGTGGAGGGAGCTCTACTAACGGACTTCAGTGGCACAGAT 903
 400 euHisAsnArgGlyGluGlnTyr.....407

904 AAGAGCTTGAGGATATGGAGGAGGACGAAGGGACGGCAAGAACCTT 948
 408LysLeuIleTrpProIleGlnGlnIlePhValArgMet 420
 seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT: AAG48655
 seq_documentation.block:
 ID AAG48655 standard; Protein; 571 AA.
 XX
 AC AAG48655;
 XX
 DT 18-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 61465.
 DE
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PR 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 990US-0121825.
 PR 05-MAR-1999; 990US-0123180.
 PR 03-MAR-1999; 990US-012548.
 PR 23-MAR-1999; 990US-0125788.
 PR 25-MAR-1999; 990US-0126264.
 PR 29-MAR-1999; 990US-0126795.
 PR 01-APR-1999; 990US-0127495.
 PR 06-APR-1999; 990US-0128234.
 PR 08-APR-1999; 990US-0129845.
 PR 16-APR-1999; 990US-0130077.
 PR 21-APR-1999; 990US-0130449.
 PR 23-APR-1999; 990US-0130510.
 PR 28-APR-1999; 990US-0131449.
 PR 30-APR-1999; 990US-0132048.
 PR 04-MAY-1999; 990US-0132484.
 PR 05-MAY-1999; 990US-0132485.
 PR 06-MAY-1999; 990US-0132486.
 PR 07-MAY-1999; 990US-0132863.
 PR 11-MAY-1999; 990US-0134256.
 PR 14-MAY-1999; 990US-0134218.
 PR 14-MAY-1999; 990US-0134219.
 PR 14-MAY-1999; 990US-0134221.
 PR 14-MAY-1999; 990US-0134370.
 PR 18-MAY-1999; 990US-0134468.
 PR 19-MAY-1999; 990US-0134941.
 PR 20-MAY-1999; 990US-0135124.
 PR 21-MAY-1999; 990US-013553.
 PR 24-MAY-1999; 990US-0135629.
 PR 25-MAY-1999; 990US-0136021.
 PR 27-MAY-1999; 990US-0136932.
 PR 28-MAY-1999; 990US-013722.
 PR 01-JUN-1999; 990US-0137528.
 PR 03-JUN-1999; 990US-0137502.
 PR 04-JUN-1999; 990US-0137724.
 PR 08-JUN-1999; 990US-0138094.
 PR 10-JUN-1999; 990US-0138540.
 PR 10-JUN-1999; 990US-0138847.
 PR 14-JUN-1999; 990US-0139119.
 PR 16-JUN-1999; 990US-0139452.
 PR 16-JUN-1999; 990US-0139453.
 PR 17-JUN-1999; 990US-0139492.

PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139464.
 PR 18-JUN-1999; 99US-0139465.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140553.
 PR 23-JUN-1999; 99US-0140554.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141442.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143524.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144225.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144332.
 PR 20-JUL-1999; 99US-0144332.
 PR 20-JUL-1999; 99US-0144332.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145244.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147088.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147304.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148561.
 PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149426.
 PR 18-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151530.
 PR 01-SEP-1999; 99US-015153070.
 PR 10-SEP-1999; 99US-0153758.
 PR 13-SEP-1999; 99US-0154418.
 PR 15-SEP-1999; 99US-015459.
 PR 16-SEP-1999; 99US-0154593.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0154486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0155458.
 PR 29-SEP-1999; 99US-0155596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157815.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158339.
 PR 13-OCT-1999; 99US-0158393.
 PR 13-OCT-1999; 99US-0158294.
 PR 14-OCT-1999; 99US-0158029.
 PR 14-OCT-1999; 99US-0159195.
 PR 14-OCT-1999; 99US-0159222.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161045.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161519.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161924.
 PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 104.00
 alignment_scores: Ratio: 0.38
 Percent_Similarity: 44.658
 alignment_block: US-09-886-400-3 x AAG48655
 Align seg 1/1 to: AAG48655 from: 1 to: 571

101	LeuProIleLeuGlu.....MetValProGluGluLeuHisPhe.....	113
135	CATAACGGCTATACTTAAAGTCTCCTCGAACAGGATATTATACACCTCG	184
114	.ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleSerGly.....	130
185	TTAAGGGGATCGAGTGACTG.....ATAGAGATA 219
130	LathLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys	146
220	ATCGGAACGGCTACACGCCAAATACCCCCCTCCGCCGTAGCAG	269
147	LysArgGinArgLeuThrLysThrMetLeuProLeu.....LeuSerGly	161
270	AGTAGAACGCCAACAGTCAGAGAGAT.....	294
161	uLeuGlyGlyIleProArgGluGlyThrLeuLeuProLysLeuLysLeuL	178
295AGGGAAAGTAAAGGAA	309
178	euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla	194
310	GAGGCTCTCGAGCTT...TCTCCCAAAGGATTCTGGCTGCCAGRCRGTGCG	356
195	GluValLeuAlaAspAsnGlyIysAspMetMetLeuProSerGlnGly	211
357	CTATGACCCGATATACTCCGTCCACTG.....	384
211	uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP	228
385AAGGACAAACGGTTATGAGTATCTATTGCCGACGGGAGGGCTTGCTT	432
228	helysAspAsnGlyIleIleLeuLeuLengluAspSerIleSerIleLeu	244
433	TTCTCAAGCTCATCTAACCTGGCATAAAGCCAATTAACCGCTCTATGCC	482
245ThrValIleLysGlyThrGly.....	251
483	ACACCTTATAAGGCCAAAAGGAAAGGGCTTAAAGCTTGCTACATCAGC	528
252LysTrpArgArgSerTerpArgTyrosIleLeuValSerAspP	265
529TATCTCTTGST.....CITC	543
265	heLeuProProSerIlysGlyGluLeuAlaLysIleLeuAspGluValLeu	281
544	AGGGAGCTTAGGAAGGGATAAAAGCTCGTTTT.....GA	578
282	GlyIleLeuArgAsnAlaIleAlaGlySerValPhenylSerIhrMetGluAs	298
579	AGGTAAGGTAACGCTAAAGGCCGTAAGCTAACATCCAAAGCGCTACCC	624
298	PGlyIysIle.....ValLysGlyLeuAlaGlyValProAspI	311
625GTTTGGGTGCCGTCGAACAGGGCTGTTAATCTCGGCATCGGAAGG	669
311	ysGlyProValLeuIleValGlyIlysIleLeuMetGlyLeuGluLeu	327
670	CTTCCTCTTATGAACTCTAAGAAACTGGCCAGCTGATAGAGGAAAGGA	719
328	GlyProMetSerGlu.....AlaPheIleLysGlyLys	338
720	CAACATTCTCTCA.....	732
339	AsnIleLeuPheArgGlyMetAlaHisProValLeuTy-SerAspAsnA	355
733TACGCCACCATATGAGTCATTGGCTAT	762
355	sPPProAlaLysAlaPheAspTyGlyAspTPrlsValPhegY.....	370
763	AGGGACATTGCAAGGCTACAGAAATGACTGTTGAGGATTATAGAGCTT	812

371AlaTyRProValThrAlaThrAsnLeuPheLysLeuL 383
 813 AGACAGGACTCTAACTCGGAACACTGCCCCCTCCCTCA.....GAGCTGA 853
 :|:||: ..::|:||: ..::|:||: ..::|:||: ..::|:
 383 uaspSerLysSerHisValLeuLeuPheProGlyAlaArgGluAlaL 400
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 854 ACCACATGGAGGGAGCTTACTAGGACTTCAGTTGGCCACGAT 903
 |:|:|: ..::|:||| :||| :||| :||| :||| :||| :||| :|||
 400 euHisAsnArgGlyGluGlnTyr..... 407
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 904 AAGAGCTGAGGATATGAGAAGGACAAGGAACTGAAAGACT 948
 |:|:|: ..::|:||| :||| :||| :||| :||| :||| :||| :|||
 408LysLeuIleIleProGluGlnGluPheValArgMet 420
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: /SIDS1/gcdata/geneseq/geneseq-emb1/AA2000.DAT:AAG16
 seq_documentation_block: Protein: 704 AA.
 ID: AAG16037 standard; Protein: 704 AA.
 XX
 AC
 XX DT 17-CCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 16523.
 XX KW Protein identification: signal transduction pathway; metabo-
 KW hybridisation assay; genetic mapping; gene expression cont-
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 PR 25-FEB-1999; 990US-0121825.
 PR 05-MAR-1999; 990US-0123180.
 PR 09-MAR-1999; 990US-0123548.
 PR 23-MAR-1999; 990US-0125788.
 PR 25-MAR-1999; 990US-0126264.
 PR 29-MAR-1999; 990US-0126785.
 PR 01-APR-1999; 990US-0127462.
 PR 06-APR-1999; 990US-0128234.
 PR 08-APR-1999; 990US-0128714.
 PR 16-APR-1999; 990US-0129845.
 PR 19-APR-1999; 990US-0130077.
 PR 21-APR-1999; 990US-0130449.
 PR 23-APR-1999; 990US-0130510.
 PR 23-APR-1999; 990US-0130891.
 PR 28-APR-1999; 990US-0131449.
 PR 30-APR-1999; 990US-0132048.
 PR 30-APR-1999; 990US-0132407.
 PR 04-MAY-1999; 990US-0132407.
 PR 05-MAY-1999; 990US-0132484.
 PR 06-MAY-1999; 990US-0132486.
 PR 07-MAY-1999; 990US-0132487.
 PR 11-MAY-1999; 990US-0132863.
 PR 14-MAY-1999; 990US-0134256.
 PR 14-MAY-1999; 990US-0134218.
 PR 14-MAY-1999; 990US-0134219.
 PR 14-MAY-1999; 990US-0134221.
 PR 14-MAY-1999; 990US-0134370.
 PR 18-MAY-1999; 990US-0133768.
 PR 19-MAY-1999; 990US-0134941.
 PR 20-MAY-1999; 990US-0135124.
 PR 21-MAY-1999; 990US-0135353.
 PR 24-MAY-1999; 990US-0135629.
 PR 25-MAY-1999; 990US-0136021.
 PR 27-MAY-1999; 990US-0136392.
 PR 01-JUN-1999; 990US-0136782.
 PR 01-JUN-1999;

PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138840.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139419.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139461.	PR	18-AUG-1999;	99US-014946.
PR	18-JUN-1999;	99US-0139462.	PR	20-AUG-1999;	99US-0147722.
PR	18-JUN-1999;	99US-0139466.	PR	20-AUG-1999;	99US-0147723.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149229.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-014902.
PR	18-JUN-1999;	99US-0139459.	PR	23-AUG-1999;	99US-014930.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149175.
PR	23-JUN-1999;	99US-0140533.	PR	25-AUG-1999;	99US-0150566.
PR	23-JUN-1999;	99US-0140534.	PR	26-AUG-1999;	99US-015084.
PR	24-JUN-1999;	99US-0140651.	PR	27-AUG-1999;	99US-0151065.
PR	24-JUN-1999;	99US-0140652.	PR	27-AUG-1999;	99US-0151066.
PR	24-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151070.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139840.	PR	23-AUG-1999;	99US-0151103.
PR	23-JUN-1999;	99US-0139841.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0139462.	PR	01-SEP-1999;	99US-0151930.
PR	18-JUN-1999;	99US-0139463.	PR	07-SEP-1999;	99US-0152316.
PR	18-JUN-1999;	99US-0139750.	PR	10-SEP-1999;	99US-0152317.
PR	18-JUN-1999;	99US-0139763.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0154038.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154179.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-015546.
PR	06-JUL-1999;	99US-0142390.	PR	13-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142870.	PR	28-SEP-1999;	99US-015656.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-015656.
PR	13-JUL-1999;	99US-0143512.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143224.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-015865.
PR	16-JUL-1999;	99US-0144086.	PR	07-OCT-1999;	99US-0158059.
PR	19-JUL-1999;	99US-0144225.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0158229.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	22-JUL-1999;	99US-0144332.	PR	21-OCT-1999;	99US-0160814.
PR	20-JUL-1999;	99US-0144884.	PR	21-OCT-1999;	99US-0156638.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145055.	PR	22-OCT-1999;	99US-0160768.
PR	26-JUL-1999;	99US-0145216.	PR	25-OCT-1999;	99US-016104.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	26-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147018.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			alignment_scores:

Quality: 104.00 Length: 365
 Percent Ratio: 0.638 Gaps: 20
 Similarity: 44.658 Percent Identity: 22.4666

alignment block:
 US-09-886-400-3 x AAG16037 ..

Align seg 1/1 to: AAC16037 from: 1 to: 704

85 ATCCAGTACGAGACACTGATAAGAGAAATTCCTTGGCTCAA 134
 ::|||::|||::|||::|||::|||::|||:
 234 LeuProIleLeuGlu...MetVal ProGluLeuHisPhe.... 246

135 CATAACGGGGTATACCTAAGGTAAAGAGAAATTCCTTGGCTCAA 184
 ::|||::|||::|||::|||::|||::|||:
 247 .ThrValP-ProTyAlaLeuSerPheIleMetGlyAspProIleysMetA 263

185 TTAAGGGGCATCGCAGTGACTG 184
 ::|||::|||::|||::|||::|||:
 263 IaThrIeuGlyIleAspGlnGlnIeuProThrIleVallysIleGlyIlys 279

220 ATCGGAACGAGCTACCGAACGAAATACTCCCOCTCCTCCCCTAGCAG 269
 ::|||::|||::|||::|||::|||:
 280 LeuArgGlnArgLeuThrLysThrMetLeuProIle....LeuSerG1 294

270 AGTAAAGCAGTCAAGTCAGAGAT..... 294

294 uLeuGlyGlyIleIleProArgGluThrIleLeuLeuTrpLeuLeuIle 311

295AGGAAGTTAACGAA 309
 ::|||:
 311 euArgSerGlyCysAlaTyRAlaAsnSerArgIleHisAlaValGlnAla 327

310 GAGCTCTTCGAGCTT...TGTCCAAAGGGATTCGGCTGCCAGGCTGCG 356
 ::|||::|||::|||::|||:
 328 GluValLeuValLeuAlaSerGlyIleAspMetLeuProSerGlnG1 344

357 CTATGCCGATAACCCGCCATACTG 384
 ::|||:
 344 uGluAlaLysArgLeuHisGlyIleLeuIleSerGlySerValArgGlySP 361

385 .AGGACAACGGTTATGAGTATCTATTGGCGACGGGGAGGGATGCTT 432
 ::|||::|||::|||::|||:
 361 heLyssAspGlyIleIsthrLeuLeuIleGluAspSerIleLeu 377

433 TTCTCAGCTCATCTCACTGGCATTAAGCCATTAAACCGCTCTATCC 482
 ::|||:
 378ThrValIleGlyIleLeuIle 384

483 ACACCPATAAGCCAAAGGAAAGGCTTAGGTACATAGC.... 528
 ::|||:
 385LysTrpArgArgSerTrpArgTrpArgTrpArgAspLeuValSerAspP 398

529TATCCTCTGCT.....CTC 543
 ::|||:
 398 heLeuProProSerLysGlyGluLeuAlaTyRAlaLeuAspGluIleLeu 414

544 AGGAGGCTTAGGAGGATAAAGGACATCAAGCGTACCC.... 624
 ::|||:
 431 pGlyIle.....ValysGlyLeuAlaGlyValroAsPL 444

625GTTGGGGCGCTGAAACAGGCTGTAAAGCTCGGCATCGGAAGG 669
 ::|||:
 444 ysGlyProValLeuLeuValGlyTyrHisNetLeuMetGlyLeuGluLeu 460

670 CTTCTCTTATGAAATCCTAAAGTGGCAGCTGATAGGGCAAGGA 719
 ::|||:
 461 GlyProMetSerGlu.....AlaPhenIleLysGluIlys.. 471

720 CAACTCTTCATA..... 732
 ::|||:
 472 .AsnIleLeuIleArgGlyMetAlaHisProValLeuTySerAspAsnA 488
 733TACGCCACCGTATAGAGTCATTCGCTATGGCTAT 762
 ::|||:
 488 SppAlaLysAlaPheAspTrpGlyAspTrpIleysValPheGly... 503
 763 AGGGACATTCGAGCTAGAATGAGGTGAGGGATTATAGGGTTAT 812
 ::|||:
 504AlaTy-ProValThAlaThrAsnLeuPhesLeuL 516
 813 AGACGAGCTCAACTCGGAACACTGCTCCCTCA.....GACCTGA 853
 ::|||:
 516 uAspIleysSerHisValLeuIleProIleGlyIalaArgGluIal 533
 854 AGCACAGTGGAGGGAGCTACTGACTTCGAGACTTCGAGTGGCACCAAGAT 903
 ::|||:
 533 eHisAsnArgGlyGluGlnTyT..... 540
 904 AGAGCTTGAGGATATGAGAGAGGAGCAAGGAAACCCGAACTT 948
 ::|||:
 541LysLeuIleTrpProGluGlnGlnGluPheValA9Met 553

seq_name: /SIDS1/gcadata/genesed/geneseqp-embl/AA2000.DAT: AAG48654

seq_documentation.block:

ID AAG48654 standard; Protein: 704 AA.
 XX
 AC AAG48654;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 61464.
 XX
 XX Protein identification: signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW terminaison sequence.

XX
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-SEP-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 990US-0121825.
 PR 05-MAR-1999; 990US-0123180.
 PR 09-MAR-1999; 990US-0123548.
 PR 23-MAR-1999; 990US-0123788.
 PR 25-MAR-1999; 990US-0126264.
 PR 29-MAR-1999; 990US-0126785.
 PR 01-APR-1999; 990US-0127462.
 PR 06-APR-1999; 990US-0128234.
 PR 08-APR-1999; 990US-0128714.
 PR 16-APR-1999; 990US-012945.
 PR 19-APR-1999; 990US-0130077.
 PR 21-APR-1999; 990US-0130449.
 PR 23-APR-1999; 990US-0130510.
 PR 23-APR-1999; 990US-0130891.
 PR 28-APR-1999; 990US-0131449.
 PR 30-APR-1999; 990US-0132048.
 PR 30-APR-1999; 990US-0132407.
 PR 04-MAY-1999; 990US-0132484.
 PR 05-MAY-1999; 990US-0132485.
 PR 06-MAY-1999; 990US-0132486.
 PR 06-MAY-1999; 990US-0132487.
 PR 07-MAY-1999; 990US-0132863.
 PR 11-MAY-1999; 990US-0134256.
 PR 14-MAY-1999; 990US-0134218.
 PR 14-MAY-1999; 990US-0134219.
 PR 14-MAY-1999; 990US-0134221.

PR	14-MAY-1999;	99US-0134370.	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	99US-0147303.
PR	04-JUN-1999;	99US-0137502.	99US-0148565.
PR	07-JUN-1999;	99US-0137724.	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	99US-014793.
PR	10-JUN-1999;	99US-0138540.	99US-0148171.
PR	10-JUN-1999;	99US-0138847.	99US-0148319.
PR	14-JUN-1999;	99US-0139119.	99US-0148341.
PR	16-JUN-1999;	99US-0139452.	99US-0148564.
PR	16-JUN-1999;	99US-0139453.	99US-0148684.
PR	17-JUN-1999;	99US-0139492.	99US-0149175.
PR	18-JUN-1999;	99US-0139454.	99US-0149175.
PR	18-JUN-1999;	99US-0139455.	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	99US-014926.
PR	18-JUN-1999;	99US-0139457.	99US-014972.
PR	18-JUN-1999;	99US-0139458.	99US-0149723.
PR	18-JUN-1999;	99US-0139492.	99US-0149929.
PR	18-JUN-1999;	99US-0139494.	99US-0149902.
PR	18-JUN-1999;	99US-0139460.	99US-0149930.
PR	18-JUN-1999;	99US-0139461.	99US-0150566.
PR	18-JUN-1999;	99US-0139462.	99US-0150884.
PR	18-JUN-1999;	99US-0139463.	99US-0152363.
PR	18-JUN-1999;	99US-0139450.	99US-015105.
PR	18-JUN-1999;	99US-0139459.	99US-015106.
PR	21-JUN-1999;	99US-0139460.	99US-0151080.
PR	21-JUN-1999;	99US-0139461.	99US-0151303.
PR	22-JUN-1999;	99US-0139462.	99US-0151438.
PR	23-JUN-1999;	99US-0139463.	99US-0151930.
PR	24-JUN-1999;	99US-0139454.	99US-0152363.
PR	24-JUN-1999;	99US-0140695.	99US-0153070.
PR	28-JUN-1999;	99US-0139763.	99US-0153070.
PR	28-JUN-1999;	99US-0139817.	99US-0153070.
PR	29-JUN-1999;	99US-0140823.	99US-0153070.
PR	29-JUN-1999;	99US-0140823.	99US-0153070.
PR	30-JUN-1999;	99US-0140991.	99US-0153070.
PR	30-JUN-1999;	99US-0141287.	99US-0153070.
PR	01-JUL-1999;	99US-0141882.	99US-0154018.
PR	01-JUL-1999;	99US-0142154.	99US-0154032.
PR	02-JUL-1999;	99US-0142055.	99US-0154139.
PR	06-JUL-1999;	99US-0142390.	99US-0154486.
PR	08-JUL-1999;	99US-0142803.	99US-0155656.
PR	09-JUL-1999;	99US-0142820.	99US-0155656.
PR	12-JUL-1999;	99US-0142927.	99US-0155656.
PR	13-JUL-1999;	99US-0143242.	99US-0155656.
PR	14-JUL-1999;	99US-014324.	99US-0155753.
PR	15-JUL-1999;	99US-0144005.	99US-0157865.
PR	16-JUL-1999;	99US-0144290.	99US-0158029.
PR	18-JUL-1999;	99US-0144685.	99US-0158232.
PR	19-JUL-1999;	99US-0144803.	99US-0158329.
PR	19-JUL-1999;	99US-0144805.	99US-0158329.
PR	19-JUL-1999;	99US-0144815.	99US-0158329.
PR	20-JUL-1999;	99US-0144827.	99US-0158331.
PR	20-JUL-1999;	99US-0144832.	99US-0158331.
PR	22-JUL-1999;	99US-0144832.	99US-0158331.
PR	21-JUL-1999;	99US-0144834.	99US-0158331.
PR	21-JUL-1999;	99US-0144833.	99US-0158331.
PR	21-JUL-1999;	99US-0144833.	99US-0158331.
PR	21-JUL-1999;	99US-0145086.	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	99US-0160741.
PR	22-JUL-1999;	99US-0145087.	99US-0160741.
PR	22-JUL-1999;	99US-0145089.	99US-0160741.
PR	22-JUL-1999;	99US-0145192.	99US-0160741.
PR	23-JUL-1999;	99US-0145145.	99US-0160981.
PR	23-JUL-1999;	99US-0145218.	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	99US-0160981.
PR	26-JUL-1999;	99US-0145216.	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	99US-0161405.

431 pglylyle..... VallysGlyLeuAlaGlyValProAspL 440
 990S-0161406.
 R 26-OCT-1999; 990S-0161359.
 R 26-OCT-1999; 990S-0161360.
 R 26-OCT-1999; 990S-0161361.
 R 28-OCT-1999; 990S-0161920.
 R 28-OCT-1999; 990S-0161992.
 R 28-OCT-1999; 990S-0161993.
 R 29-OCT-1999; 990S-0162142.
 R 29-OCT-1999;
 alignment_scores:
 Quality: 104.00 Length: 365
 Ratio: 0.638 Gaps: 20
 Percent Similarity: 44.658 Percent Identity: 22.466

Align seg 1/1 to: AAG48654 from: 1 to: 704
 US-09-886-400-3 x AAG48654 ..

Alignment block:
 85 ATCCAGTATCGAGACGTAAAGAGAAATTCTTTGGCTAA 134
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 234 LeuProIleIeuGlu...MetValProGluLeuHisPhe..... 246
 135 CATAACGGCTATACTTAAAGTTCACTCCTCCGAAGGATTTATGACCTCG 184
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 247 .ThrValProTyraLeuSerPhenylAlaNetGlyAspProIleLysMetA 263
 185 TTAAGGGGCATCAGACTCGAGTCAGCTGACTG 219
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 263 IthrLeuGlyIleAspArgGlnLeuProThrGlyIvalLysIleGluLys 279
 220 ATCGAACGGCTACAGCACGCAAATCTCCCTCTCCCGTTAGGAG 269
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:
 280 LeuArgGlnArgLeuThrLysThrMetLeuProIeu.....LeuSerG1 294
 270 AGTAGAAAGCAACAAGTCAGAGAT..... 294
 ::|||:|||:|||:|||:|||:|||:
 294 uLeuGlyGlyIleIleProArgGluThrLeuLeuTrpLeuLysLeuL 311
 295AGGGAAAGTTAACGAA 309
 ::|||:
 311 euArgSerGlyCysAlaTyraLalaSerArgLeuAlaGlnAla 327
 310 GAGCTCTGAGCT...TCTCCAAAGGGATTCTGGCTGCCAGAGCTCG 356
 ::|||:|||:|||:|||:|||:|||:|||:
 328 GlyuValLeuValLeuAlaSerGlyIysAspMetLeuProSerGlnG1 344
 357 CTATGACCGATAATCCGCCACTG..... 384
 ::|||:|||:|||:|||:
 344 uGluAlaLysArgLeuHisIleLeuLeuLysAsnCysSerValArgCysP 361
 385 ..AAGCACAAAGGTATGACTATCATGCCGAGGGAGGCCATGTT 432
 ::|||:|||:|||:|||:|||:|||:
 361 heLyAspArgNlyFisThrLeuLeuGluAspSerIleSerLeuL 377
 433 TTCTCAGCTCATCTCAACTGGGATAAACGCCATTAAACCGCTCTATCC 482
 ::|||:|||:
 378ThrValIleGlyIleGly..... 384
 483 ACACCTTATAAGGCCAAAGGGAAAGGCGCTTAAAGGTACATCAGC... 528
 ::|||:|||:|||:|||:|||:
 385LysTyraArgSerTriArgTyAspLeuValSerAspP 398
 529TATCTCCTGGT.....CTC 543
 ::|||:|||:
 398 heLeuProProSerLysGlyGluLeuAlaIleAspGluValLeu 414
 ::|||:
 544 AGGGCCTTAAAGCCTAAAGGTCTGTTTT.....GA 578
 ::|||:|||:
 415 GlyPheLeuArgAsnAlaValGlySerIlePheSerThrMetGluLAs 431
 579 AGGTAAAGTAAAGCTAAAGGACTAAAGACATCGAAAGCTGCGTACCC... 624

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AA16166
 seq_documentation_block:
 ID AAG16036 standard; Protein: AAG16036;
 XX DE 17-OCT-2000 (first entry)
 AC XX Arabidopsis thaliana protein fragment SEQ ID NO: 16522.
 AC XX DE Protein identification; signal transduction pathway; metabolic pathways; gene expression; gene mapping; gene hybridisation assay; genetic mapping; gene termination sequence.
 AC XX KW Arabidopsis thaliana.
 AC XX EP1033405-A2.
 AC XX PD 06-SEP-2000.
 AC XX PF 25-FEB-2000; 2000EP-0301439.
 AC XX PR 25-FEB-1999; 990S-0121825.
 AC XX PR 05-MAR-1999; 990S-013180.
 AC XX PR 09-MAR-1999; 990S-0123548.
 AC XX PR 23-MAR-1999; 990S-0125788.
 AC XX PR 25-MAR-1999; 990S-0126264.
 AC XX PR 29-MAR-1999; 990S-0127462.
 AC XX PR 01-APR-1999; 990S-0127462.
 AC XX PR 06-APR-1999; 990S-0128234.
 AC XX PR 08-APR-1999; 990S-0128744.
 AC XX PR 16-APR-1999; 990S-0129845.
 AC XX PR 19-APR-1999; 990S-0130077.
 AC XX PR 21-APR-1999; 990S-0130449.
 AC XX PR 23-APR-1999; 990S-0130510.
 AC XX PR 23-APR-1999; 990S-0130891.
 AC XX PR 28-APR-1999; 990S-0131449.
 AC XX PR 30-APR-1999; 990S-0132048.

PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-014508.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-014508.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-014508.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-014508.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-014519.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-014514.
PR	11-MAY-1999;	99US-0134216.	PR	23-JUL-1999;	99US-014521.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-014524.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-014526.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-014591.
PR	14-MAY-1999;	99US-0134310.	PR	27-JUL-1999;	99US-014518.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-014591.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-014591.
PR	20-MAY-1999;	99US-015124.	PR	02-AUG-1999;	99US-014638.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-014638.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-014730.
PR	08-JUN-1999;	99US-0136392.	PR	03-AUG-1999;	99US-014740.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-014730.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-014719.
PR	03-JUN-1999;	99US-0137588.	PR	05-AUG-1999;	99US-014726.
PR	04-JUN-1999;	99US-0137592.	PR	06-AUG-1999;	99US-014730.
PR	07-JUN-1999;	99US-0137774.	PR	06-AUG-1999;	99US-014741.
PR	08-JUN-1999;	99US-0138044.	PR	07-AUG-1999;	99US-014743.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-014793.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-014730.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-014831.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-014839.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-014856.
PR	17-JUN-1999;	99US-0139492.	PR	14-AUG-1999;	99US-014856.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-014936.
PR	18-JUN-1999;	99US-0139445.	PR	17-AUG-1999;	99US-014937.
PR	18-JUN-1999;	99US-0139446.	PR	18-AUG-1999;	99US-014942.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-014972.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-014973.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-014992.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-014992.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-015056.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-015088.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-015106.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-015108.
PR	22-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-015130.
PR	23-JUN-1999;	99US-0140091.	PR	31-AUG-1999;	99US-015138.
PR	24-JUN-1999;	99US-0140333.	PR	01-SEP-1999;	99US-015193.
PR	24-JUN-1999;	99US-0140334.	PR	07-SEP-1999;	99US-015205.
PR	24-JUN-1999;	99US-0140895.	PR	10-SEP-1999;	99US-015307.
PR	28-JUN-1999;	99US-0140833.	PR	13-SEP-1999;	99US-015358.
PR	03-JUL-1999;	99US-0142930.	PR	15-SEP-1999;	99US-015401.
PR	12-JUL-1999;	99US-0142977.	PR	16-SEP-1999;	99US-015403.
PR	13-JUL-1999;	99US-0142887.	PR	20-SEP-1999;	99US-015477.
PR	01-JUL-1999;	99US-0141842.	PR	22-SEP-1999;	99US-015513.
PR	01-JUL-1999;	99US-0142254.	PR	05-OCT-1999;	99US-015773.
PR	02-JUL-1999;	99US-0142055.	PR	06-OCT-1999;	99US-015865.
PR	06-JUL-1999;	99US-0142350.	PR	23-SEP-1999;	99US-015516.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-015519.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-015832.
PR	19-JUL-1999;	99US-0144290.	PR	12-OCT-1999;	99US-015839.
PR	19-JUL-1999;	99US-0144325.	PR	29-SEP-1999;	99US-015856.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-015893.
PR	19-JUL-1999;	99US-0144364.	PR	13-OCT-1999;	99US-015894.
PR	19-JUL-1999;	99US-0144405.	PR	14-OCT-1999;	99US-015935.
PR	16-JUL-1999;	99US-0144405.	PR	14-OCT-1999;	99US-015936.
PR	19-JUL-1999;	99US-0144434.	PR	14-OCT-1999;	99US-015937.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-015938.
PR	20-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-015939.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-015963.
PR	21-JUL-1999;	99US-0145096.	PR	18-OCT-1999;	99US-016074.
PR	21-JUL-1999;	99US-0145096.	PR	21-OCT-1999;	99US-016074.

XX 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSEQ INC.
 XX
 PI Drmanac RT., Liu C., Tang YT.
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS82457.
 XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
 PT
 PT
 PT
 PT
 XX Claim 20; SEQ ID NO 48629; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences, (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The CC
 CC Polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving or (III). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. ABC0010-ABG3037 represent novel human diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
 XX Sequence 936 AA;
 SQ
 alignment_scores:
 Quality: 103.50 Length: 312
 Percent Ratio: 0.778 Gaps: 19
 Similarity: 42.628 Percent Identity: 22.436
 alignment_block:
 US-09-886-400-3 x ABG18270 ..
 Align seg 1/1 to: ABG18270 from: 1 to: 936
 379 ATATCGAAGGACAAACGGTTATGAGTATCTATTCCGGACGGGGAGGCAT 428
 195 uletyr.....ValGluserluvalGluasnglyLeuproThr 209
 183 LeuLeuGluAspAsnGlyIyp 510
 429 GCTTTTCAGTTCACTCAACTCCGATAAAGCCAATAAACGCCCT 478
 195 uletyr.....ValGluserluvalGluasnglyLeuproThr 209
 479 ATCCACACCTATAAGGCCAAAGGGAAAG 510
 209 alProAlaAsnItpSerLeuHisArgGluIysvalAlaAlaPheAlaGly 225
 511GCCTTTAGGTACATCAGCATCCTTGATCTCTGTC 543
 226 AlaGluProTyroHsItpSerLeuAsnIsthrLeuGlyPheProArgValGly 242
 544AGGGAGCTTGGAGGGCATAAAGCTGGTTTTGAAGGTAAGG 586

```

OM of: US-09-886-400-3 to: SPTRMBL_19:*
out_format : pfs

Date: Jun 11, 2002 10:39 PM

About: Results were produced by the GenCore software, version
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-NODEL=frame+npb,model -DEV=x1h
-0=/cnr2/17USPTO.spool/us9886400/runat_11062002-150916-174
-DB=SPTRMBL_19 -OFM=fastan -SUFFIX=rstd -GAP0P=12.000
-GAPEXT=4.000 -NMATCH=0.100 -LOOPCL=0.000 -LOOPENT=0.000
-QAPOP=4.500 -QGAPEXT=0.050 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -START=1 -MATRIX=blossom2
-DELOP=6.000 -DELEXT=-7.000 -LIST=-45 -DOCNAME=THR_SCORE=pct
-TRANS=human40.cdb -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -NODE=LOCAL -OUTFMT=pfs
-NUSER=US-09-886-400 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US-09-886-400 -ICPUS=6 -ICPDI=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPXY -WAIT -THREADS=7

Search information block:
Query: US-09-886-400-3
Query Length: 1095
Database: SPTRMBL_19:*
Database sequences: 562222
Database length: 1729995929
Search time (sec): 61.230000

```

sp.bacteriap:O25577	+	97.50	129.66	6.54	1085	1
sp.archaeap:O8:239	-	97.00	136.18	6.67	396	1
sp.bacteriap:O66973	+	97.00	136.18	6.67	66973	1
sp.archaeap:O7BF8	+	97.00	135.79	6.19	97BF8	1
sp.bacteriap:O32582	+	97.00	134.45	6.29	610	1

```

Command line parameters:
-MODEL=frame.n2p model
-O=/cgn1.2.1/USPO-POOL
-DB=SPIREMBL.19 -QFMT=f-a
-GAPEXT=4.000 -MINPATT=0
-GAPPOP=4.500 -OGAPEXT=7
-FGAP=6.000 -FGAPEXT=7
-DELW=6.000 -DELEXT=7
-TRANS=human-cdA -LIST
-THRMAX=100 -THR MIN=0
-NOM=ext -HRSIZE=500
-USER=US09886600@CGN1.1
-DEV_TIMEOUT=120 -WARN_T

Search information block
  Search: US-09-086-400-3
  Query: length: 1095
  Database: SPRINTMBL_19:*
Database sequences: 56222
Database length: 17299494
Search time (sec): 61.2391

```

Documentation	
Q9C105	Pyrococcus furiosus
Q96XV6	Pyrococcus horikoshii
P74610	Synechocystis sp. (str. PCC 7002)
Q975F3	Thermoplasma volcanium
Q973E0	Clostridium acetobutylicum
Q977NO	Clostridium butyricum
Q976ZD	Clostridium acetobutylicum
Q977NO	Sulfolobus tokodaii
Q50094	Pyrococcus horikoshii
Q9K4D4	Bacillus halodurans
Q9Y294	Schizosaccharomyces pombe
Q9h191	Thermoplasma acidophilum
Q93046	Archaeoglobus fulgidus
Q9V0M7	Pyrococcus abyssi
Q973E0	Hypothalamic tuberculum
Q83337	Treponema pallidum
Q977NO	Sulfolobus solfataricus
Q976ZD	Sulfolobus solfataricus
Q9bkv1	Leishmania major
Q94317	Schizosaccharomyces pombe
Q996XV6	Schizosaccharomyces pombe
Q93218	Mycobacterium tuberculosis
Q977NO	Sulfolobus solfataricus
Q9zn2	Acaribiosis thalhana
Q9mbD0	Pyrifolia (Japan)
Q93131	Oncorhynchus mykiss
Q93132	Oncorhynchus mykiss
Q94233	Chlamydia trachomatis
Q97442	Clostridium acetobutylicum
Q93182	Treponema pallidum
Q9451	Schizosaccharomyces pombe
Q88138	Mus musculus (mouse)
Q17585	Gaenorhabditis elegans
Q97442	Acetilicus hysudris
Q94076	Leishmania major
Q91075	Leishmania major
Q92z2	Yersinia pestis (trans-
Q9zqf1	Rickettsia prowazekii
Q97442	Acaribiosis thalhana
Q98PQ2	Mycoplasma thalhana
Q98PQ2	Mycoplasma pulmonis

```

--> - Quality: 1532.00 Length: 363
      Ratio: 4.441 Gaps: 0
Percent Similarity: 95.041 Percent Identity: 79.063

alignment_block:
US-09-186-400-3 x Q9HB55 . .

Align seg 1/1 to: Q9HB55 from: 1 to: 364

  1 TTTGAGGGCGCTCGTCCTTACGGCAAACCTCCAGTAGCCGAATCCCAAA
  1 MetGluAlaLeuAlaIpheHisGlyAspLeuGlnTyRAlaGluIleProLys
  51 GAGGAAATCCCCAAGGTCTATAGAGGGCATACATCCCAGTCATGAGAGA
  17 sSerGluIleProLysValleGluIlysAlaItyrPheProThrIleSerG
  101 CACTGATAAAAGAGAAATTCCGTTGGCACAATAACGGGTATACC
  34 IleLeuIleArgArgGluIleProPheIleLeuAsnIleThrGlyTyrSer
  151 TTAAAGTTCCTCCGAAAGGATATTAGACCTCTGTTAAAGGGGATCTGCC
  51 LeuSerPheLeuIleProLysAspLeuIleAlaIleGluIleGlyTieGlu
  201 GAGTGACCTGGATAGAGATAATCGGAAGGAGCTAACAGGCACCCAATACTCC
  67 uSerGlyLeuIleGluIleLeuGlyThrSerTyrThrHisAlaIleLeuP
  251 CCCCTCTCCGCTAGAGAGAGAAACACAAAGTTCAGAGGATAGGGAA
  84 roLeuLeuProLeuSerArgValGluIlaGinIleAspGaspArgGlu
  301 GTTAAAGGAAAGACCTCTGAGCTTTCGAAAGGGATCTGGCTGGCAGAA
  101 ValysGluAsnIleLeuGluIvalAspGlyIphePheLeuProGlu
  351 GCTGGCCTATGACCCGATAATCCCTGGCATACTGAGGACACGGTTATG
  117 uLeuAlaTrpAspProIleIleProAlaIleLeuIleAspGlnAsnTyrG
  401 AGTATCTATTGCGGACCGGAGCGATGCTTCTGAGTCATCTCAAC

```


601 GTCAGAACATCGAAGCCGTACCGTTGGGCGCTGAAACGGCTGT 650
 201 VallysAspIleGluAlaIleProValIrrpValSerLeAlaIall 217
 651 AATGCTCGCATGGAAAGCTCCTCTATGATACTCAAGAAATGGCGA 700
 217 eMetLeuGlyAlaGlyArgPheProLeuMetSerProLysArgValAlaA 234
 701 GCTGGATAGGACAAGGACACATTCAGGGACCAGTATAG 750
 234 snTrpIleLysGlyLysAspGluIleLeuLeuLeuLeuLeuLeu 250
 751 TTICATGGCTATAGGGCACATTGAGGTACAGAAATGTTGAGGGAT 800
 251 PheLeuGlyIleGlyArgSerIleAlaGlyHisLysIleAlaIleSerGlyLe 267
 801 ATTAGACGTTATAGCAAGCTTAACCTGGAAACTGGTCCCTCAGACC 850
 267 umetGluValLeuAsnGluIleGlyGlyGlueucyIeuproArgAspI 284
 851 TGAAGCACAGTGAAAGGGAGCMCTACTACCGAACCTGAGTTGGGCCACCA 900
 284 learginHisAsnGlyArgLeuIrrpIearginThrSerSerIrrpIalro 300
 901 GATAAGGCTTAGGATGGAGAGACGAGGAGCAGGAGCAGACCTAA 950
 301 AspLysserIleArgIleTriplysGluAspGluGlyAsnAlaArgLeuAs 317
 951 TATGCTCTCTACATAATGAGGGGAACTCCCTTTAGCCGAGAACAA 1000
 317 nMetLeuThrIrrpIearginThrSerAspIglyGluPheAlaIpheLeuAlaIunrs 334
 1001 GCGATGAAAGGGATGGAGGCCCTCCCTGAGGGAGCTGGATGCCCCTC 1050
 334 erAspAsparGlyIleTriplysGluProIgLuProIgLuProIgLu 350
 1051 CGGGCGATATAAACATGGAGGGTCAAATGGGAA 1089
 351 LysIleAlerryIrrpIearginGluAsnGlyLyS 363
 name: sp_bacteriAp:P74630
 documentation_block:
 P74630 PRELIMINARY; PRT; 529 AA.
 MEDLINE#-97051201; Published=880523;
 01-FEB-1997 (TREMBLrel. 02, Created)
 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 01-JUN-2001 (TREMBLrel. 12, Last annotation update)
 HYPOTHETICAL 62.1 kDa PROBIN.
 SLL0735.
 Synochocystis sp. (strain PCC 6803).
 Bacteria; Cyanobacteria; Chroococcales; Synechocystis .
 NCBI_TaxID=11448; "
 SEQUENCE FROM N.A.
 MEDLINE#-97051201; Published=880523;
 Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakam
 Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.
 Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K.,
 Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Ya
 tabata S.;
 Sequence analysis of the unicellular cyanobac
 terium, Synochocystis sp. strain PCC6803. II. Sequence determina
 tion of the entire genome and assignment of potential protein-coding re
 gions. DNA Res. 10:109-136 (1998).
 EMBL: D90916; BAA18743.1; -.
 InterPro: IPR00303; DUF200.
 Pfam: PF02651; DUF200; 1.
 Hypothetical Protein; Complete protein.
 SEQUENCE 529 AA; 4C1A45048A784E30 CRC64;

`alignment_scores`:

510 GCGCTTGTAGGTACATCAGCTATCCTCTGGTCAGGGAGCTTAGGAAGG 559
 :::|||||::|||::|||:|||:
 273 aGlutYrGluTyrylElSsProTyrylMetProAspGlyInArgIysA 290
 ||||::|||:|||:|||:|||:
 560 CGATAAACCTCGTTTT.....GRAAGTAAGGAAACG 591
 ||||::|||:|||:|||:
 290 snIleGlyIleLysTyryHisLysIleThrSerArgAspGlyIleLeuSer 306
 ||||::|||:|||:
 592 CTAAAGGCAGTCAGAACATCGAAGCCGTACCGTTRGGTGCCCCGAA 641
 ||||::|||:|||:
 307 GluIysAlaTrpTyAsp.....ProTyrrpAlaLysGluLy 319
 ||||::|||:
 642 CACGGCTGTA.....
 ||||::|||:
 319 sAlaAlaGluAlaSerAsnPhemMetTyrasnArgGlnGlnGlnValG 336

652ATGCTCGGATC ..GGAGGGCTCCCTTAAGAATCTAGAAA 693
 ::|||:|||:|||:|||:
 336 lYHisLeuSerGlyIleMetGlyArgProProLeuValSerProTyR 352
 ||||::|||:
 694 GTGGCGGC.....
 ||||::|||:
 353 AspAlaGluIePheGlyIleTrpTrpTrpGlyProTrpHeIleAs 369
 ||||::|||:
 712GACAGGCAACATT..... 726
 ||||::|||:
 369 pTyriLeuPheArgLysSerTrpHeAspGlnAspThrPheGluMetThrH 386
 ||||::|||:
 727CITCTATATAGGCCACCGATAcAGTCATT..... 756
 ||||::|||:
 386 lsLeuAlaAspTyryIleArgGlyIysProHisIglnValcysAspGPro 402
 ||||::|||:
 757GGCTATAGGGACATGGCTAC..... 780
 ||||::|||:
 403 SerGlnSerSerItpGlyIleTrpHisGlyIleTrpIleAsnAs 419
 ||||::|||:
 781AGATGA 787
 ||||::|||:
 419 pThrAsnAlaTrpIleTyProHisLeuIlysAlaIgluArgIetI 436
 ||||::|||:
 788 GTGTGAGGGATTNTAGGGTTATAGGAGCTCAACTCGGAACCTGTC 837
 ::|||:|||:|||:|||:|||:
 436 leGluIeuserHisArgGluAlaValAlaAspGluIeIgluGluIys. 450
 ||||::|||:
 838 CCTCCCTCAAGCTGAAGACAGTRGAAGGACGGCTCRAFTIA...CGCAC 884
 ||||::|||:
 451AlaIeAsnGlnAlaAlaIgluIleLeuIleAlaGlnSe 464
 ||||::|||:
 885 TTCGAGTTGGCA 897
 ||||::|||:
 464 rSerAspTrpAla 468

seq_name: sp_archeap:Q97BM4

seq_documentation_block:

ID Q97BM4 PRELIMINARY; PRT; 378 AA.

AC Q97BM4

DT 01-OCT-2001 (TREMBREL, 18, Created)

DT 01-OCT-2001 (TREMBREL, 18, Last sequence update)

DT ALPHA-AMYLASE.

DE TVG0421416.

GN Thermoplasma volcanium.

OS Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmae;

OC Thermoplasma.

NCBI_TaxID=50339;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAN=GSS1 / DSM 41299 / JCM 9571;

RX MEDLINE=20570466; PubMed=11121031;

RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S., Kawamoto T., Yamada K., Makino H., Suzuki M., NunoShiba T., Yamanoto Y., Aramaki H., Makino K., Suzuki M.,

RT "Archaeal adaptation to higher temperatures revealed by genomic sequence of Thermoplasma volcanoium";

RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).

DR EMBL; AF000592; BAB59573.1; -.

KW Complete proteome.

SQ SEQUENCE 378 AA; 44957 MW; 49FCCF63BEE6B3A7 CRC64;

alignment_scores:

Quality: 149.50 Length: 313

Percent Similarity: 46.965 Gaps: 11

Percent Identity: 21.086

alignment_block:

US-09-886-400-3 x Q97BM4 ..

Align seg 1/1 to: Q97BM4 from: 1 to: 378

64 AAGGTCTAGAGAGGCCATACICCAGTCATCGAGACACTGATAAAGA 113
 ||||::|||:|||:|||:|||:
 42 ArgValAlaGluArgSerTyrlleProAlaIhrArgAsnLeuMetGluTy 58
 ||||::|||:
 114 AGAAATTCTTGGCTCAAAATAAGCTACGGC ..
 ||||::|||:
 58 rGlyIleIlysSerSerPheSerIleThrGlyIhrAlaValGluGinAlaL 75
 ||||::|||:
 145 ...TATACCTAAAGTCCICCCGAAAGGATAATTAGACCTCGTTAAA 189
 ||||::|||:
 75 eumMetTyAsnTrpIlys.....ValueAspAlaIleAsp 86
 ||||::|||:
 190 GGGGCCAICGGAGTAGCTGATAGAGATAATCGGAAACGAGCTACCGCA 239
 ||||::|||:
 87 AspTyryValysSerGlyLeucyIleLeuGluIleUtrTyTyRhi 103
 ||||::|||:
 240 CGCATACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCAAAGGGATTG 289
 ||||::|||:
 103 sSerIeuAlaSerIleItpAspTyraspGluIphelysArgGlnValAspM 120
 ||||::|||:
 290 GAGATGGGAAGTTAGGAAGAACCTCTCGGCTTCTCCAAAGGGATTG 339
 ||||::|||:
 120 ethIaSArgAspIleMetIysArgIleIleAspIleAspValProysValPhe 136
 ||||::|||:
 340 TGCTGCCAGAGCTCGGCTATGACCTCTCCATACTGAGGAA 389
 ||||::|||:
 137 ArgAsnThnGluIuleItyraspAspArgIleAlaIgluMetVallySar 153
 ||||::|||:
 390 CAACGGTTATGAGTATCPATTCGCCAGGGAGGCCATGCTTTCAG 439
 ||||::|||:
 153 gMetGlyIpheHsSerIleIleIhrGlyIlys..... 163
 ||||::|||:
 440 CTCATCTCAACTGGCGATAAAACCAATTAAACCGCTCTATCCACACCTT 489
 ||||::|||:
 164ThrIspSerIleIleValAspHisSerSerOAntry..... 175
 ||||::|||:
 490 ATAAAGGCCOAAGGGAAAGCCCTTAGGTAAcATGAGC.....TA 530
 ||||::|||:
 176ArgTyrlAlaserProSerGlyLe 183
 ||||::|||:
 531 TCTCTCTGGCTCAGGGAGCTTGAAGGCGATAAAAGCTGTTTGAAG 580
 ||||::|||:
 183 uAsnIeutyIleArgAsnTyrlAlMetSerAspAsnIleSerPheArgP 200
 ||||::|||:
 581 GTAAGGTAACGCTAAAGGAGCTAAAGACATCGAAGCGTACCGTTGG 630
 ||||::|||:
 200 heSerAsnThr.....LysTrpIysAspTyPhePro..... 209
 ||||::|||:
 631 GTGGCGCTGAAACAGGCGCTGTAATGCTCGGCATGGAAAGGCTCTCTTAT 680
 ||||::|||:
 210Le 210

681 GATCCTANGAAAGTGGCAGCTGGATAGGAGAACAGAACATCTTC 730
 ||||::|||:
 210 utrIaAlaAspIlyStyAlaIstPheAsnGluSerGluGlyAspMetV 227

271 heGluLeuProMetGluTyrIleLysProTyrlleAsnGlu.....Asn 285
 829 GAACTGTGCCTCCCTCAGAGCTGAAGCAC.....AG 860
 286 GlyIleIearleAspThrGlyIleIleGlyAsnSe 302
 861 TGGAGGGAG...CTCTACTTAGGACTTCGAGT.....TGGG 895
 302 rglycIuIysGlyIleIetTrasnargIuAsnAlaMetIlysvaltrpG 319
 896 CACCGATAGAGTGTGAGGATGAGAGAGCAGGGAAC..GCA 942
 319 luHisAlaserHisPheAlaSerCysArgHisAspGinIleAsnAlaA 335
 943 AGACTATAATG..... 954
 336 AlaAlaAsnMetAspIlysproProIleThrCysProTyrlAspIhrG1 352
 954 954
 352 uLeuTyrglyHistpTrpPheGluGlyProAspPheIleAsnAlaPheI 369
 955 955
 369 learIlySsAlaGluasPPrpThrseryGluLeuIleThrPrrohr 385
 982 GCCCTTTAGCGAGAACAGCGATGCAAGGGATGGAGGCCCTCCCTGA 1031
 386 GlutTyIeIuIysAsnSerMetAlgIcySerProSerProPro... 401
 1032 GAGGGGGCTGGATGCCTCGGGGGATATAACGATTGGGGGTGAA 1081
 462 462
 1082 ATGGGAA 1089
 407 sngIAsp 409

seq_name: sp_archeap:Q973T0

seq_documentation_block

ID Q973T0 PRELIMINARY; PRT; 443 AA.

AC Q973T0

DT 01-DEC-2001 (TREMBREL; 19, Created)

DT 01-DEC-2001 (TREMBREL; 19, Last sequence update)

DT 01-DEC-2001 (TREMBREL; 19, Last annotation update)

DE HYPOTHETICAL PROTEIN S10817.

ST017.

GN Sulfobulus tokodai.

OS Crenarchaeota; Sulfobolales; Sulfobacae; Sulfobolus.

NCBI_TaxID=111955;

[1]

RN SEQUENCE FROM N.A.

RP STRAN=JCM 10345 / 7;

RC PubMed=11572479;

RX Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Yamagishi A., Ohshima T., Kikuchi H.; "Complete genome sequence of an aerobic thermoacidophilic Crenarchaeon, Sulfobolus tokodaii strain7"; DNA Res. 8:123-140 (2001).

RW EMBL; AP000983; BAB5830_1; Hypothetical protein; Complete proteome; SEQUENCE 443 AA; 53218 MW; 497087264B03C4 CRC64;

KW Hypothetical protein; Complete proteome; SEQUENCE 443 AA; Length: 343

SQ Quality: 142.50 Gaps: 15 Percent Identity: 23.032

alignment_scores:

Quality: 142.50 Length: 343

Ratio: 0.902 Gaps: 15 Percent Identity: 23.032

Percent Similarity: 46.064

865AGGGAGCTCACTTACGG..... 882
 278 luPhelysLeuProAgluValTyrLysGluTyttyraspGluIleVal 294
 883ACTCGAGTGGCA.....CGAGATAAGAC... 909
 295 IleAspSerThrValSerTrPAlaAspIleAsnLysAspGluserTr 311
 910TGAGGATATGAGAGGAGCAAGGAAGCAGACITAAGCTTAAGATGAA 949
 311 pheGlyAsnIleMetGlnTrpAlaTyrAspGluMetValArgArgThrg 328
 950 ATATGCCTCCATACATTATGAGGGCGAA 978
 328 IuMetLeuAlaLysGluAlaGlyGlyGlu 337
 seq_name: sp_archeap_Q972N0
 seq_documentation_block:
 ID Q972N0
 AC Q972N0
 DT 01-DEC-2001 (TREMBrel_19, Created)
 DT 01-DEC-2001 (TREMBrel_19, Last sequence update)
 DT 01-DEC-2001 (TREMBrel_19, Last annotation update)
 DE HYPOTHETICAL PROTEIN ST1102.
 GN ST1102.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfobacae; Sulfolobus.
 RN N1
 RP SEQUENCE FROM N A.
 STRAIN_ID 10545 / 7;
 RX PubMed:1157479;
 RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
 Naito Y., Nishijima K., Otsuka R., Nakazawa H., Takamaya M., Kato Y.,
 Yochizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 Oshima T., Kikuchi H.;
 "Complete genome sequence of an aerobic thermoacidophilic
 Crenarchaeon, Sulfolobus tokodaii strain 7.";
 RL DNA Res. 8:123-140(2001).
 EMBL; AB00985; BAB6615; 1.
 HYPOTHETICAL PROTEIN: Complete proteome.
 SEQUENCE 895 AA; 101064 MW; 6BCT/CD8380DDFBEC CRC64;

alignment_scores:
 Quality: 136.50 Length: 410
 Ratio: 0.669 Gaps: 20
 Percent Similarity: 49.756 Identity: 20.732

alignment_block:
 US-09-886-400-3 x Q972N0 ..

Align seg 1/1 to: Q972N0 from: 1 to: 895

7 GCCTCGTCCTTCAGGCCAACCTCAGTAGTCGGAAATCCCCAAAGAGCGA 56
 377 SerLeuValGlyAlaTyrIleLeuLeuIleLysGlnPheAs 393
 57 ATGCCAAAGGTCTATAGAAAGGAAATACATCCCAAGTCATC 96
 393 nvalSerValThrIleAsp.....ProValLeuLeuIleUrySint 408
 97 .GASACACTGATAAGAGAA.....ATTCCTTGG 129
 408 rPGluThrIleLeuHisGluIleAsnAlaThrPheThrSerAsnPheGly 424
 130 CTCACATAAGC.....GGTATACCTTAAGTCTCT 161
 425 ValAsnValSerHisAspIleGluAlaValAsnTyrThrLeu..... AlaleuG1 696

438GAGCAGAGGAAACGCA 56
 439AsnLeutYrLysThrLeuIleLysGluIlylsv 450
 212 TAGAGATAATGGAAACGAGTACAGCACCCAACTACTCCCTCCTC ... 258
 450 alGluValLeuThrValProPheThrProLeuIleLeu 466
 259CCGTTAGCATAGATAAGCAAGTCAAGATAGATAGGAA 299
 467 GluAspGlyTyrTrpSerAspValLeuAlaGlnLysGluMetGlyGluAs 483
 300 AGTTAACGAAAGGCTTTCGAGCTTCTCCAAGGATCTGGGTGCCAG 349
 483 metThrHisGluVaPheGlyVaTrpAlaAspGlyIleTrpThrProG 500
 350 AGCTCGGCTATGACCCGATATCCCTGCCATACTGAAAGCAACGGTTAT 339
 500 LuMetAlaPheAspPmTAspIeuValGlyLeutYrAsnGluSerGlyIle 516
 400 GAGTACTATCGCGACGCGGGAGGGATGCTT...TCCTCAGCTCATCT 446
 517 SerPheThrIleLeuAspGlnGlnAlaPheLeuProTyrValThrLeuVa 533
 447 CRACTGGCGATAAAAGCCAAATTAAACCCTGCTATCCACACCTTATAAANGG 496
 533 IasGlySerLeuAspProAspGlnIroPheIleValGlu..... 546
 497 CCCAAAGGAAAGGCCCTTATGATCATCAGCTATCTCTGGTCAGG 546
 547Asn 547
 547 GAGCTTAAAGCGATAAAGCTGTTTTGAAAGGTAAGSTARAAGCTAAA 596
 548 AsnLeuGlyGlnThrIleLeuVallePheArg...AsnThrThrLeuSe 563
 597 :.GGAGTCAAAGACATCGAAGGCTACCGGTTGGTGCG 637
 563 rAsnGluPheGlyPheSerGlnSerProGlnLeuThrAla. 579
 638 TGAACACGGCTSTAATGCTCGGCATCGGAAGGCTTCCTCTATGATCT 687
 580GlnGluLeuIleGlnGlnLeuAlaGluIleTyrMetAsnPro 594
 688 AGAAAGTGGCAGCTGATAGGGACAAGGACAACATCTCTATACGG 737
 595 GlyGlyValValThrValAlaLeuAspGlyIleuAsnProLeuIlePheAs 611
 738 CACGGATATAAGTICATGGCTATAGGGACATGG 772
 611 nProThrThrGlyProSerAspLeutYrAla...TyglnAlaLeus 627
 773 CAGGCTACAGAATGAGTGTGGGGATTATAGGGTTATAGGAGCT 822
 627 ergIutYrGlnGlySer.....TrpLeuValThrIleThr 638
 823 ACTCGGAACTGTCGCTCCCTCAGRCGTCAGCAGTGAAGGGAGCT 872
 639 AlAsnSerGluAlaIle.....AlaThrHisYsProThrSerIleIleTh 653
 873 CTACTTAGGACTTCGAGTGGCACCAAAGATAAGACTTGAGATATGG 922
 653 rAsnLeuProValAsnSerTrp.....AspLeuAsnLeuAsnTyrIpa 668
 923 GAGAGCAGAGGAAACGCA 942
 668 snAsnGlyTyrIleGlyLysIleLeuIleLysGluIleTrPheAsnValSerLeuAla 684
 943 AGACITATAATGGCTTCATACATAATAGGGCCGAACTCGCCCTTTAGC 992
 685 ArgGluIutYrLeuIleAlaTyrIleLeu.....AlaLeuG1 696

129 eargglueLegrlnaspGluglyTyrvalluvalliehrserAlaaLat 146
 236 CGACGGCAATACTCCGCCCTCCGTTAGCAGA...GTAGAACACAA 282
 146 hrhislytyrLeuProLeuLeuLysArgAspAlaAlaLysLysAlaGin 162
 283 GTCAGAGAGATAGGAAGCTTAAAGAGCTCTGAGGTTCGAGCTTCGAAA 332
 163 LeuleuAsnGlyLysValyGluLysTyrGlyArgLysProAR 179
 333 GGATTCGGCCAGACTCGGATAGCCGATA..... 369
 179 qglylterplleuplqGlycysAlaTyrArgProAspGlyLeutPLys 196
 370ATCCCCGCCTACTG 384
 196 erProSerThrGlyGluvalLysTrPArgLysGlyIleGluHisLeu 212
 385 AAGGCAACGGTTATGACTCATTCGCTGGGGAGGGATGTTT 434
 213 LysLysPheGlyIleGluLysTrPheValGly..... 223
 435 CTCAGTCACTCGCATTAACGCTAAACGCCATTAAACCGCTCTAAC 484
 224 ...SerHistLeuIleAspLys ..GlyProValSerLeuArgTgIgLyA 238
 485 ACCTTAAAGGCCAAAGGAAAGCCGTTAGGTACATGAGCATCTC 534
 238 snIleLeuProAla ...LysThrLysArgSerThrLeuArgProTrPhe 253
 535 CTGGCTCTAGGGACTTGGAAAGCGTAAAGCTCTTGTGAAGTAA 584
 254LeuIysAsnGlyIleAlaValIheAlaArgsnAR 265
 585 GGTAAAGCTAAAGGAGTCAAAGCATGAAAGCTTGTGGTGG 634
 265 gGluThr..... 636
 274 1a 274

seq_name: sp_bacteriap:09KD04

seq_documentation_block:

ID	Q9KD04	PRIMINARY;	PRY;	923 AA.
AC	Q9KD04			
DT	01-OCT-2000	(TREMBUREL, 15,	Created)	
DR	01-OCT-2000 (TREMBUREL, 15,	Last sequence update)		
DT	01-JUN-2001 (TREMBUREL, 17,	Last annotation update)		
DE	BH1415 PROTEIN.			
GN	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/clostridium group;			
OX	NCBII_TaxID=86655;			
RN				
RP				
RC				
STRAIN	C-125	/ JCM 9153;		
STRAIN-C				
CREATED				
CREATOR				
MEDLINE	20512562; PubMed=11058132;			
RX				
RA	Takami H., Nakasono K., Takeki Y., Maeno G., Sasaki R., Maeda F., Hirama C., Nakamura Y., Ogawara N., Kubara S., Horikoshi K;			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with <i>Bacillus subtilis</i> Nucleic Acids Res. 28:4317-4331(2000); DR			
DR	AP001512; BAB05124; -			
DR	InterPro: IPR03803; DUF200.			
DR	InterPro: IPR001296; Glycos_transf_1.			
DR	PF02651; DUF000; 1.			
DR	PF00534; Glycos_transf_1; 1.			
DR	Complete proteome.			
SPORTNAME	923 AA. 106638 MW. RC787D64F603553 CRC64;			

94	ATCGAGACATGATTAAGAAATCCTTTGGGTCAAATAACGGG	143
2236	ValGluThrValLeuLysHisGlnMet...TrpLeuAsnHisthr..	250
1144	CTATACCTTAAGTCTCCGGAAGGATTATAGAACGCTCTAAAGGG	193
2251PheGluGlyIleAsnLeuLeuGly.	262
1194	GCAATGGCAGTGTACTGTATGAGATAATCGGAAACGACTAACGACGCA	243
2263AsnGlyAsnValGluValThrValAlaProTyroThrHisPro	276
2244	ATATCCGCCCTCCCG.....CTTAGAAGAGTAGAAAGCACAA	281
2277	IleGlyProIleAsnAspPheGlyTrpTrpGluAspPheAspAlaGly	293
2282	AGTCAGAGAGATGGAAAGTAAAGGAAAGACTCTTC.....GAGC	322
2293	nValLysIysAlaasnGluLeuLysGluTyreLysGlyAlaGlyLysv	310
3223	TTCCTCAAAGGGATTGCGCAGAGCTCGCTATGACCGAAATC	372
3310	alThrProLysGlyGlyTrpAlaAlaGluSerAlaLeuAsnAspLysth	326
3373	CCTGCCATACTGAGGAACAACGGTTAGGTATCATWTCGCCGAGGG..	420
3327	LeuGluIleLeuAlaGluAsnGlyTrpLysTPValMetThrAspGinLe	343
421GAGGGCATGCCTTCTCACTGCATCAACTCGGGATAAAGC	463
3343	uvallLeugluLysLeuGlyValProLystrIleGluSerIrrTyryLysp	360
4464	CA.....	465
3360	roTrpValAlaGlnPheGlyAspLysIlePheProArgAsn	376
4465	465
3377	HisAspLeuSerAspArgValGlyPheArgTyralaGlyMetAsnGlnTy	393
4466ATTAACCGCTCTATCACCTTATAAAAGCCCAAAGGGAAA	508
3393	rAspAlaValLysAsnPheValGluLeuLeuLysIleGluLysGlnA	410
5509	AGCCCTTCTAGGTACATCAGCTATCCTCTGGTCTAGGGACCTTGAAG	558
4410	sTrpAspGlySerLeuValTyrvallIeThLeuAspGlyGluAsnPro	426
5559	GCGATAAAAGCTCTTGGTAAAGTGAACTAACGCGCTGAA.....AA	605
4427	TrpIoluHistyrPhePheAspGlyLysLeuPheLeuGluLeuTyar	443
6060	AGCATCGGAAGCGCTTACCGCTTGGTGGCCGTGAAACGGCTGTA.....	651
4443	gGInLeuGluIleGluLysSglYLeuIeArgHrValThrProS	460
6652ATGCTCGGC.....ATCGGAAGCTTCTCTTATG	681
4460	ergIutyryIleGluMetPheGlyAspLysAlaLysLeuThrProLys	476
6682	AATCTAAGAATGGCCAGCTGAGGAGAACGAAATATCTCTCT	731
4477	MetMetLysArgLeuAspPheThrThrGluAspAsnValAlaLeu	493
7732	ATAGGCAACGAAATAGAGTTCTGGCATATGGACATTCGAGCTACA	781
4493	uIValAlaLysThrIle.....	GlyGluL 501
7782	GAATGAGTTGTTGAAAGGATTATAGAGTTATAGAGGACTCAACTCGGAA	831
5001	eUTrAspMetValGlyValThrClu.....	509

SQ	SEQUENCE	602 AA;	70219 MW;	326B1198D54D43E CRC64
406	CTATTGCCGAGGGGAGGTGATTTCTAGCTCATCTCAACTCGGC 455			
154	.. . GluglnGluGlyLrhtryProleuProGluAlaGlnMetAspArgPh 169			
456	GATAAAACCAATTAAACCGCTCATCACACCTTATAAGCCAAAGGG 505			
169	etLeuArgNetaArgMetArgProGlyTrpGluSerIleGluGluMetG 186			
506	AAAAGCCTTAAAGGTACATCAGCTAT..... 531			
186	LutLeuArgArgArgIleSerTrpArgLysAspAspProThrIleAsp 202			
532CCTCTGGCTCAGGGAGCTTAGGAAGGGAT 563			
203	ValGlu-ProValValSerLeuGluLrhPheArgAqIleGlnAspAlaVa 219			
564	AAGCTGTTTGAAAGTAAAGCTAAAGCAGTAAGACATCG 613			
219	IgluAlaValValTrpValAspLysSerIleLeuLysTrpSerGluLeuV 236			
614	AACCCGTACCCCTTGGTCCCGTAAACGGCTAAAGCCGATC 663			
236	aLargAla.....ThrArgGluHisGluLeuValGluLeuIleSer 249			
664GGAGGGCTTCCTTGTAAATCTCTGAAAGTGGCAGCTG 704			
250	SerProArgGlyIleLeuAlaLeuLysLeuAlaArgAlaLeuAlava 266			
705	GATAGAGACAGGGACACATCTT.....CTATAGGCCACCGATA 745			
266	lMetAspPolyArgAspPheValIleProAspAspValAlav 283			
746	TAGAGTCATTGGCTATAGGCCATTGAGCTTACAGAAATGAGTGTGAG 795			
283	aIgluAlaLeuAlaHisArgValIleLeuLysPheGluTrpAlaValGlu 299			
796	GGATTATA.....GAGGTATAGACGAGCTCAACTCGGAACTGTGCCT 839			
300	GlyLeuArgAlaGluGluValValGluIleLeuAsnSerValArgva 316			
840	TCCTCTAGAGCTGAAGCAGCTGAAGGGAG 870			
316	IPro.....LysItyGluAlaGlnGlu 323			
seq_name: sp_archeap_Q9VOM7				
seq_documentation_block:				
ID	Q9VOM7	PRELIMINARY;	PRT;	602 AA.
AC	Q9VOM7;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	HYPOTHETICAL PROTEIN.			
GN	PA1857.			
OS	Pyrococcus abyssi.			
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.			
OX				
RN				
RP				
RT	Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution"; Submitted "2001-07-1999" to the EMBL/GenBank/DDBJ databases.			
URL				
DR				
DR	InterPro; IPR003803; DUF290.			
DR	InterPro; IPR003882; HHH_.			
DR	InterPro; IPR003851; DUF200.			
DR	InterPro; IPR003883; HHH_.			
DR	PF02651; DUF200_1.			
DR	SMART; SM00278; HhH1_1.			
DR	PROSITE; PS00049; ALDOKETO_REDUCTASE_2; UNKNOWN_1.			
DR	HYPOTHETICAL Protein; Complete proteome.			
KW				
seq_documentation_block:				
ID	O83377;	PRELIMINARY;	PRT;	526 AA.
AC	O83377;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CONSERVED HYPOTHETICAL PROTEIN.			
GN	TP0158.			
OS	Treponema pallidum.			
OC	Treponema; Spirochaetales; Spirochaetaceae; Treponema.			
RN	[1]			
RP	SEQUENCE N. A.			
STRAIN	NICHOLS;			
MEDLINE	98333770; PubMed=9665876;			
RX				
RA	Fraser C.M., Morris S.J., Weinstock G.M., White O., Su			
RA	Dodson R., Gruim M., Hickey E.K., Clayton R., Ketcham			
RA	Soerergren E., Hardham J.M., McLeod M.P., Salzberg S.,			

Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback J.A., McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garber J.A., Hatch B., Horst A., Kettler R., Koenig S., Sandusky M., Weismann J., Smith A., Venter J.C.
 "Complete genome sequence of *Treponema pallidum*, the syphilis spirochete." *Nature* 399:314-318 (1999).
 EMBL: A001215; AAC65444.1; - .
 TIGR: TP0358; - .
 InterPro: IPR003803; DUF200.
 Pfam: PF02651; DUF200; 1.
 Complete Proteome:
 SEQUENCE 525 AA; 59699 MW;
 Q D5FOC32CADE6C41A CRC64;

```

Alignment_scores :          Quality: 111.00      Length: 216
                                         Ratio: 1.047      Caps: 7
Percent Similarity: 49.074      Percent Identity: 19.907

Alignment_block:
US-09-8886-400-3 x 083377    .

Align seg 1/1 to: 083377 from: 1 to: 526
10 AlaphValleAspIysAsileupPropheValargGlyAlaGlyAlaSer 26
43 .....ATCCAAAGGCCGAAATCCCGTCACTCCAGGAACTCCAGTCGGAA..... 42
26 rSerIleUAlaGluSerArgPheHeLeuGluLisSerTyrTyrTyrL 43
86 TCCCATCATC.....GAGACACTGATAAAGAAAGAAATTCTCTTT 126
43 euProleUleArgLeuCysGluIleGluArgGluArgValProphe 59
147 GGCTCACATA..... 138
60 AsnIleSerIleAlaIleGlyProvalleUcysGluMetIleAlaasnAR 76
138 ..... 138
76 gValleuMetAspArgTyrArgGalaLeuAspIleLeuIleGluPheG 93
138 ..... 138
93 lyGluArgGluAlaIleArgLeuArgAsnSerLeuGlnGluArgValGln 109
139 ACGGGTATACCTTAAGTTCTC.....CCGAAGGATATTAGACCT 182
110 AlaGluAlaAlaLeuIleArgSerIleArgSerHisArgAspTyrPheAsphi 126
183 CGTTAACGGGCCATCCGGTAGCTGACCTG..... 210
126 SCysAspGlyAlaAlaLeuIleGluArgIleAsnHisPhePheArgThrGlyS 143
211 ..ATAGGATAAACTGGAACGACTACAGCAGCAAATACTCCCCTC... 255
143 erIleGluLeuAlaIleThrAlaValAsnCysPheLeuProPhyTyr 159
256 ....CTCCCGCTTAGCAGACTAGCAGACTAGCAGAACAAAGTTAGAGAGATGGGA 299
160 GlnAspProducIle.....IleSerAlaGlnIle**MetGlyLeuI 175
300 AGTTAACGGAGGCCTCTTCGAGCTTCTCCAAAGGGATTCTGGCTGCCAG 349
175 eAsnTyArgLyShisPheSerIleProArgLypheTyrLeuIroG 192
350 ACCTCGCTATGACCCATAATCCCTGCCATACTGAGGACAACGGTAT 399
192 IleGluGlyTyrAlaPrcA+AlaGluIleGluArgGlnIleSerIleGlyPhe 208

```

```

  400 GAGTATCTATTGCCGACGGGAGGGATGCTTTCTAGCTCATCTC 447
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
  209 SerThrHisLeuGluThrSerPheLeuPheGlyThrArgVal 224

```

```

  400 GAGTATCTATTGCCGACGGGAGGGATGCTTTCTAGCTCATCTC 447
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
  209 SerThrHisLeuGluThrSerPheLeuPheGlyThrArgVal 224

```

```

DM of: US-09-886-400-3 to: PIR_71: * out_format : pfs
Date: Jun 11, 2002 10:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+np2,model1 -DEV=x1h
-Q=Cgn2.1/USPOT_spool/US0886400/runat_11062002-150914-17315/app_query.fasta_1.1169
-DB=PIR_71 -FORMAT=fasta
-SUFFIX=tpr -GAPOP=12.000 -GAEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPXT=0.000 -OGAPOP=4.500
-QGAPOP=0.050 -XGAPOP=0.000 -FGAPOP=6.000
-QGAPEXT=7.0000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELET=7.0000 -START=1 -MATRIX=Ilossum62 -TRANS=human40_cdi
-LIST=45 -DOCOLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-AALIGN=15 -MCODE=LOCAL -OUTENT=pis -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=20000000000 -USER=IS09886400 @@CGN1.1.121
-NCPB=6 -TCPB=3 -LONGHC=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

```


R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Se M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu DNA Res. 5: 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A;Reference number: A71000; MULD:98344137
A;Accession: G71241
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-633 <KAW>
A;Cross-references: GB:AP000001; NID:93236128; PID:BA929262-1; PID:93256579
A;Experimental source: strain On3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa C;Genetics:
A;Gene: PH0193
C;Superfamily: Dictyoglomus thermophilum amylase A

alignment_scores:
Quality: 161.50 Length: 414
Ratio: 0.820 Gaps: 25
Percent Similarity: 47.585 Percent Identity: 23.430

alignment_block:
US-09-886-400-3 x G71241 .. .

Align seg 1/1 to: G71241 from: 1 to: 633

22 GGCAACCTCCAGATGCCAAATCCAAAAGGGAAATCCCAAAGGCAT 71
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
17 GlyAsnPhiEuiPile.....
17 GlyAsnPhiEuiPile.....

72 AGAGAGGCAATACTCCAGTCAGAACACTGATTAAGAGAAATTCC 121
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
27 rGluIysAlaTyrArgProPhiLeuGluThrIle.....GluGluTyrP 42

122 CT.....TTGGGTCAATAQGGTATACCTTAAGTCTC 162
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
42 roAMetIysValAlaHisIleGlyValLeuValIleIutTrIeu 58

163 CCGANG.....GATATTAGACCTCGTAAAGGGGGATCGGAG 203
:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
59 GluArgAsnArgProGluTyrIleAspLeuLeuIysSerLeuIleSly 75

204 TGACCTGATAGAGATACTGGAAAGGAGTCACAGCACCAAACTCCCC 253
:::|||:|||:|||:|||:|||:|||:|||:|||:|||:
75 sGlyGlnValGluLeuValAlaGlyPhetYrGluProIleLeuValA 92

254 TCCCTCCGCCTTACAGAGTAGAAACCACAAAGTCAAGAGCATAGGAAGTT 303
:::|||:|||:|||:|||:|||:|||:|||:|||:
92 IaiIlePro.....
97 GluGluAsp 97

304 AAGGGAAAGCTCTCGAGCTTCT 327
:::|||:|||:|||:|||:|||:|||:|||:
98 ArgValGluInleIysLeuSerLysGlyTrpAlaArgGlyMetGlyTy 114

328 ...COAAGGGATTCTGCTGCGAGACTCGCCTATGACCGATAATCC 373
:::|||:|||:|||:|||:|||:|||:|||:|||:
114 rGluIlaArgArgIleLeuTrpLeuThrGluArgValTrpIeuProGluLeu 131

374 CTGCATACTGAAAGCAACGGTATAGGTATCTATGCCCACGGGAG 423
:::|||:|||:|||:|||:|||:|||:|||:
131 allylStIleArgIlaAlaGlyIleGluTyValIleLeuAsp..... 145

424 GCGATGCTTTCTGAGTCATCTCAACTGGGAGCTCTAAGGCTT 470
:::|||:|||:|||:|||:|||:
146AspTyrHisPheMetSerAlaGlyLeuSerLysGluIgl 158

471 ACCGGCTCATCCACACTTATAAGGCCAAAGGAAAGGCCCTTATGGT 520
:::|||:|||:|||:|||:|||:
158 uLeuPhiTerPProTyrTrpGluAsnGlyGlyGluAlaIleValW 175

seq_documentation_block:
Probable alpha amylase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998
C;Accession: G71241
521 AC.....ATCAGCTATCCTCTGCTAGTCAGGAGCTTACGG 555

seq_name: PIR2:G71241

A Gene: amyA
 C Superfamily: Dictyoglomus thermophilum amylase A
 C Keys: Glycosidase; hydrolase
 F # 2-686/Product: amylase A #status experimental <MAT>
 alignment_scores:
 US-09-885-400-3 x ALDYAT ..
 Align seg 1/1 to: ALDYAT from: 1 to: 686
 alignment_block:
 22 GCAACCTCCAGTAGGCCAAATCCAAAAGGGAAATCCCCAAA
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 18 GlyAspPheAspPheVal.....
 72 AGAGAAGGCTATACTCCACTCATCGAGACTGATTAAAGAA.
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 28 rgluMetSerTyrLysProLeuIleAsnPhePhePhelysHSIP
 119 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 119 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 45 heProIleAsnVal11HisPheSerGlyPheLeuLeuTerPheu
 169 GATATTAATAGACCTCGTTAAAGGGGCATCGCGAG
 62 AsnHilProGluTyrrheGluLysLeuWtSilenMetAlaGilar
 210 GATAGAGATAATCGGAACGAGCTCACGGACGAAATACTCCGCC
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 78 nineGluPheValAspSerGlyLysPheTrpIleLeuProIle
 260 GCGTTCAGGTAGGCAAGGTTCAGAGATAAGGAAATTT
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 95 roAspLysAspLysValGlnGlnLysLysLeuAspLysLystyr
 310 GAGCCTCTGAGCTTCCTCCAAAGGCAATCTGGTGCCAGAGCT
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 112 AspLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly
 360 TGACCCGATAATCCCTGCCATACGAAAGGACAACGGGTATGAGT
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 128 PheIleProIleValLysTyrIleIleGluAAGlyIleGluT
 410 TCGCCGACGGGAGGCATGCTTCTCAGCTCATCTCAAACCTG
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:
 145 alvalaAsp...AspAlaHisPheSerValGlyLeu....
 460 ANGCCAATTAAACCGTCTATCCACACCTTATAAGGCCAAAG
 :|||:|||:|||:|||:|||:|||:|||:|||:
 157
 :|||:|||:|||:|||:|||:|||:|||:|||:
 510 GCGCTTCTAGGTACATCGATCTCTGCTCTGAGGAGCTTA
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 160 pLeuPheGlyIyr....TyrLeuMetGluGluGlnGlyTyrL
 560 CGATAAACGCTCTTGTAAAGGTAAACGCTAAAGGAGCTC
 |||||:::|||||:::|||||:::|||||:::|||||:::
 175 la.....ValPheProLeuSerMetLysLeuArgTyrIleLeu
 610 ATGGAAGCCGTACCGTTGGTGGCCGIGAAACACGGCTGTAA
 :|||:|||:|||:|||:|||:|||:|||:|||:
 186
 :|||:|||:|||:|||:|||:|||:|||:
 660 CATGGAAAGGCTTCTCTCTATGATTCCTAAGAAAGTGGCGAGCT
 :|||:|||:|||:|||:|||:|||:|||:
 187TleProPheAlaAspProGluGluThrIleThrIleThr
 709GKGCAAGGACAAATTCTCTATACGCG

310 GAGCCTTCAGCTCTCCAAAGGATTTGGCTGCCAACAGCTGCCATA 359
 111 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 132 GluIlePheGlyPheLysAlaLysValIleArgAnThrGluLeuIleLty 148
 360 TGACCGATAATCCGTCCATACTGAAGGAAACCGTATGAGTATCT 409
 111 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 148 ArgAsnArgLleAlaLysIleAllysAPlsLeuIlyPhyLysAlaIleP 165
 410 TGCCTGACGGG... GAGGGCATGCTTTCTCAGCCTCATCAACCGGG 456
 111 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 165 heRnGluGlyIleGluLysIleIeu..... 173
 457 ATAAGCCAAATTAAACCGCTCTATCACACCTATAAAAGCCCCAAAGGG 506
 111 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 174
 507 AAAGGCCTTAGGTACATCAGC..... TATCTCTTGGTCAGGG 547
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 177 rProAsnTyrlLeutyrGlnSerPrAspGlyMetIysIleIeuLeuArgA 194
 548 AGCTTAGGAGGGATAAACCTCGTTTTGAAAGGAAGGTAAACGGCTAAAG 597
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 194 snyTyArgLeuSerAspPheGlyPheArg..... Phe 205
 598 GCAGTCAAAGCATGAAACGGCTAACGGCTGTAAACGGCTGTAACTGGCTCTC 676
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 206 SerAlaArgAspPrAspGlnTyrProLeuThrAlaAspLysTyAlaIal 222
 627 TGGGGGCCGTGAAACGGCTGTAAACGGCTGTAAACGGCTGTAACTGGCTCTC 676
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 222 eTrpIleAlaSerIleProLeuGlyIleValIleAsnIle..... T 235
 677 TTATGAACTCCAAAGAAGTGGCAGC..... TGCAATAGGACAAAGGAC 720
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 235 yrMeAspTyrlGlutIhrPheGlyGluHisIlePheGluIu..... 248
 721 AACATCTCTTCTATACGGCACGGATA... GAGTCATGGCTATAGG... 765
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 249 ThrGlyIlePheGluIleLeuArgTyrlLeuPr 259
 766 ... GACATGGAGGTAC... AGATAGAGTGTGGATTATAGGAGG 808
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 259 oIleGluIleAlaLysIleGluIleGluValAlaValSerGluIv 276
 809 TTATGAGGAGCTCAACTCGGAAACTGGCTTCCCTCAGAGCTGAAAGCAC 858
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 276 alvaAlaPheLeuGluUpProArg..... 283
 859 AGTGGAGGGAGCTACTACGG..... ACTTCGAGTTGGCCA... 897
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 284 GlyGluIleIlyrvalHisGluPheAlaTrileSerTralaas 298
 898 .CCAGATAAGGCTCTAGGATATGGAGAGGAGGACGAAACGGZAGAC 946
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 298 pThrGluIargAspValSerAlaIlePhe..... GlyAsn... LysM 311
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 947 TTATATGCTGCTTCAATAATGAGGGGGACTCC... CTTTAGCC 993
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 311 estGlnArgIleSerIlePheGluIysLysAspIleIlySphIleIlys 327
 994 GAGAAAGCAGCATGCCAGGGATGGAGGCCCTCCCTGAGAGGGCTGGA 104
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 328 GluAsnSerAlaLysIle..... LysLysLeuAs 337
 1044 TGCCCTCGGGCGATAATAACGATGGAGG 1074
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 337 nLysIleAspGluIleIlyrLysMetIlyrLys 347

C; Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
 C; Accession: A49512
 R; Laierman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen, C.E.
 J. Biol. Chem., 266, 24402-24407, 1993
 A; Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus. Q
 A; Reference number: A49512; MVID:94043280
 A; Accession: A49512
 A; Status: preliminary
 A; Molecule type: DNA
 A; Residues: 1-649 <LAD>
 C; References: GB:L22346; NID:9347939; PIDN:AAA72035.1; PID:9347940
 C; Genetics:
 A; Gene: amyA
 A; Start codon: GTG
 C; Function:
 A; Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
 A; Pathway: Glycogen/starch degradation
 C; Superfamily: Dityoglomus thermophilum amylase A
 C; Keywords: glycosidase; hydrolase; polysaccharide degradation

alignment_scores:
 Quality: 141.50 Length: 391
 Ratio: 0.773 Gaps: 23
 Similarity: 46.803 Percent Identity: 22.762

alignment_block:
 US-09-886-400-3 x A49512 ..
 Align seg 1/1 to: A49512 from: 1 to: 649

73	GAGAGGCATACATCCAGTCATCAGACACTGTTAAAGAAATTCC	122
29	GlutylS cysteinyrrProLeuGluThrLeu.....GluGlyTyrPr	43
123	T.....TTGGCTCAACATAACGGCPATACCTTAAGTCTCC	163
43	O ASNMetLySvalAlaLeHishSerGlyProLeuIleGluTrpLeuG	60
164	CG.....AGGATATTATAGACCTCGTAAAGGGCATCGGAGT	204
60	InAspAsnArgProGlyTyryrLeuLeuArgSerLeuValysArg	76
205	GACCTGATAGAGATAATGGAAAGGAGTACAGCACGAAATACTCCCCT	254
77	GlyInvaGluLevalvaAAGlyProValLeuLeuLase	93
255	CCTCCGCTTAGGAGAGTAGAACACAAGTAAAGGAGATAGGGAGTTA	304
93	rilebro.....LysGluAspA9GlyGlnIleArgLeuMetI	107
305	AGGA.....GACTCTTCGACCTTCCTCAAAAGGATTCTGCRGCCA	348
107	YsGluTrpAlaLySserIleGlyPheAspAlaArgGlyValTrpLeuThr	123
349	GAGCTGCCTATGCCGATAATCCCTCAATGAGGAAACGCTTA	398
124	GluArgValTrpGnProluLeuIvalystrLeuIysGluSerGlyI	140
399	TGAGTATCTATTCCGCAGGGGGATGCCCTTCAGCTCATCTCA	448
140	eAspIlyValIleAsp.....AspTyrHisPheM	151
449	ACTGGCGATAAAsCCAATTAAACGCTCTAT...CCACACCTTAAAG	495
151	etSerAlaGlyLeuSerIlysGluIleLeuTy-TrpProt-TyrThrGlu	167
496	GCCAAAGGAA.....AAGCGTTAGTGA	521
168	AspGlyGlyLuvalleAlaValPheProIleAspGluLySLeuArg.	183
522	CATAGCTATCCTGAGGTAGGAAAGGCTTAAGGAAAGCTAAAGTC.	570

seq_documentation_block:
 hypothetical protein PH1386 - Pyrococcus horikoshii
 C; Species: Pyrococcus horikoshii
 C; Date: 14-Aug-1998 #text_change 20-Jun-2000
 C; Accession: D71011
 R; Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Ogi, M.; Ohfuki, Y.; Funanashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogura, DNA Res., 5, 55-76, 1998
 A; Reference number: A71000; MVID:98344137
 A; Reference: 1-560 <KAN>
 A; Molecule type: DNA
 A; Residues: 1-560
 A; Status: preliminary
 A; Sequence: nucleic acid sequence not shown; translation not shown
 A; Cross-references: GB:AP000006; NID:93226133; PID:BR30492.1; PID:93257809
 A; Experimental source: strain Or3
 A; Note: this accession replaces an interim accession for a sequence replaced by GenB.

274 1a 274

318	CGAGCTTCTCAAGGATTCTGCTGCCAGCTGCTTGTAGACCCCA	367
172	eGlyLysProLeuGlyPheTerPheLeuProGlyUratAlaPhesTerProG	189
368	TAATCCCTGCTCATACTGAACGACAACGGTTATGAGATCTATCGCGCAT	417
189	IyvaAspArgIleLeuPheGluGlyIleArgTyThrPheValAsp	205
418	GGGGAGGCGATGCTTTTC.....TCAGCTCATCTCAACTGGCGAT	458
206	GluHisAlaIvalLeuThrAlaAspProThrProIlysGlySerSer..	221
459	AAAGGCAATTAAACGGCTCTAT...CCACAC.....CTTAAAGG	496
222	[: : : : : : : : : : : : :	235
497	CCCAAGGGAAAAGGCCTTAGGTCATCAGCTCATCCTGGTCAGG	546
235	rgHisthrGluLeuSerAlaLysValTrpSerSerThrLeuGly.....	249
547	GAGCTTAGGAAAGGGATAAAGCTCTTTTGAAAGCTTAACCTAA	596
250	[:: : : : : : : : : : : : :	257
257	gGluPheTerArgAspIleAlaTyrasPArgGluLysPheTerPheLeuGly	274
597	GGCA...GTCAAAGGACATCGAAGGGTACCGCTTGG.....	630
274	roHistvalHISAspGlyIleArgIleAspThrGly.....LeuLys	288
661	ATCGGAGGCTCCCTATGAATCTAAAGAA.....GRCGCGAG	701
289	TyrHisArgIleThrGlyHisthrGluGluLysAspLeuTyraValArgG	305
702	CTGGTAGAGGACAAG.....	717
305	utrpAlaGluLysArgValGlnGluHisAlaAsnHisPheIleGlyAlaI	322
718	[:: : : : : : : : : : : : :	726
322	LeHistHisGluIleAspGlnHisGlyGlnAsnIpheProProTyrvai	338
727	CTTCCTTACGGCACCGATATAGAGTCATGGCTATGGCACATTGAGG	776
339	MetValIthrProPheAspAlaGluIlePheDelyHisthrTrpPheGluG	355
777	CTACACAATGAGTGTGAGGATTAGAGTT.....ATAG	814
355	yProGlnTrp...IleGluIalaLeuTyrGluGlnGlyAlaAspArgVal	371
815	ACGAGCTCAACTGGAACTGTGGCCTTCCTCAGAGTGAAGCACAGTGGAA	864
371	erTerIleTerProGluLeuTerIleGlnIargHistryGln...AspPhe	386
865	AGGGAGCTCTACTTACGGACTTCGAGTGGCACAGATAAGAGCTTGAG	914
387	GlnThrAlaHisValSerPheSerTerIrrTerIrrTerIrrTerIrr	403
915	GATATGGAGAACGGCAAGGGAACCAACACTTATAATGCTGTCCTACA	964
403	svalPheAsnAspHis...AspAlaIatrPmetTyArgHistyTrHisA	419
965	ATATGAGGGCAACMCGCCCTTTA.....	990
419	rgMetGluLysAspIleAlaLysValAlaMetTerProGlnProThr	435
991	[:: : : : : : : : : : : : :	1024
436	ValLeuGluLysGlnAlaIleGlnGnMetValAlaGluItrp.....	449
1026	CCCTGAGAGGGAGGCGATGCTTGGGGGATATATAAGATTGG	1071

```

450 ::::::::::::::::::::MetLeuAlaValSerAspPrrp 457
seq_name: pir2:A75207
seq_documentation_block:
  seq_documentation_block:
    amyTopullulanase PAB0122 - Pyrococcus abyssi (strain Orsay)
    C;Species: Pyrococcus abyssi
    C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_chu
    C;Accession: A75207
    R;Anonymous, Genoscope
    submitted to the EMBL Data Library, July 1999
    A;Description: Pyrococcus abyssi genome sequence: insights
    A;Reference number: A75001
    A;Status: preliminary
    A;Molecule type: DNA
    A;Residues: 1-1362 <KAR>
    A;Cross-references: GB:AJ248283; GB:AL096636; NID:95457433;
    A;Experimental source: strain Orsay
    C;Genetics:
    A;Gene: apu; PAB0122

alignment_scores:
  Quality: 124.00 Length: 381
  Ratio: 0.756 Gaps: 16
  Percent Similarity: 43.045 Percent Identity: 21.522

alignment_block:
  alignment_scores:
    Quality: 124.00 Length: 381
    Ratio: 0.756 Gaps: 16
    Percent Similarity: 43.045 Percent Identity: 21.522
  alignment_block:
    US-09-886-400-3 x A75207 .
    Align seg 1/1 to: A75207 from: 1 to: 1362
      94 ATCGAGACACTGATTAAAGAAGAAAATTCTTGTGGCTCAACATAAACGGG
      236 ValGluIleValLeuLysHisGlnMet...TrpLeuLeuIleHistThr...
      144 CTATACCTTAAGTCCCTCCGAAGGATATTAGACCTCTAAAGGGG
      251 .....PheGluGluLysGluLysIleAsnIleLeuIleGluGluGluGlu
      194 GCATCGGATGACCTGAGAGATAATCGGAAGAGTCACGCCACCGA
      263 .....AsnGlyAsnValGluValIthrValProTyRThrHisPro
      244 ATACTCCCCCTCCCTCCCGC .....CTTAGAGAGTAGAAAGGCCACA
      277 TleGlyProIleLeuAsnAspPheGlyTrpGlyCluAspPheGlyIle
      282 AGTTCAGAGAGATAAGGGAAAGTTAAGGAAGAGCTTC.....GAGC
      293 nvalLysLysAlaLysGluIeuntryLysGluIyLeuGlyAlaGlyLysV
      . TTTCCTCAAGCCGATTCCTGCCTGCAGCTTCGCTCATGACCGATATTC
      310 althPrLysGlyLysIleAlaGluSerAlaLeuAsnAspLysthr
      . CCTGCATACATGAAAGCACACGGTTATGAGTATCTATTCGCGACGGG...
      327 LeuGluIleLeuAlaGluIleGlyLysIleGluSerIleGluSerIle
      421 .....GAGCGCATGCTTTCTCAGCTCATCAACCTGCGGATAAAGC
      343 uValLeuGluLysLeuGlyValProLysThrIleGluSerIleGluSerIle
      464 CA .....
      | |
      360 rotrpvalAlaGlnPheGlyAspLysIleIleIleIleIleIleIleIle
      465 .....
      377 HisAspLeuSerAspArgAlaGlyPheArgTyRAlaGlyMetAsnGlnTy

```

```

466 .....ATTAACCGCTCATCCACACCTTATAAGGCCAAAGGGAAA 508
393 rAspAlaLysAsnPheValGluLeuLysIleGlnLySgIna 410
509 AGCCTTAACTGGTACATCAGCTATCCTGGCTCAGGGACCTTGAAG 558
410 snYrAspGlySerIeuValTyrrValIeuThrLeuPheGlyLysAsnPro 426
559 GCGATAAGCTGAGGTAAAGGTAAACCTTAAGGCAGTC...AA 605
427 TrpGluHisItyrProPheaspGlyLysLeuPheLeuGluIleLeuTyR 443
606 AGACATCGAACCGTACCCGTTGGTGCCGTTAACCGGGCTGA.... 651
443 gGlnLeuGluGluLeuGlnLysGlyLeuIleArgThrValhPROs 460
652 .....ATGCTGGC.....ATCGGAAGGCTTCCCTCTATG 681
460 ergLutYrIleGluMetPheGlyLysAlaIysLeuThrProLys 476
682 ATTCCTAAGAAAGTGGCGAGCTGGATAAGGCAAGAACATTCTCT 731
477 MetMethylsargLeuAspPheThrThrGluAspAsnValAlaLeuLe 493
732 ATACGGCACCCATPAGAGTCATGGCTATAGGCAATGAGGTCTACA 781
493 UlysAlaLysThrLei.....GlyGluL 501
782 GAATGACTGTGAGGGATTATTAGGGTTATAGAGGTATAGAGCTCAACTCGGAA 831
501 eutYrAspMetValGlyValIthrGlu..... 509
832 CTGTCGCTTCCTCAGAGCTGAAGCACAGTGGAAAGGGAGCTCATCTAGC 881
510 ..... 514
514 GluMetLnnTrPrP 514
882 GACTTGCACTGGCACCAAGCATAAAGCTTGAGGATATGGAGGAGSACG 931
514 oGluSerSerTrPle...AspGlyThrIleSerThrTrP.....I 527
932 AAGGGAAACGCAAGCTTAATATGGCTGTCTCAATATAAGGGCGAACTC 981
527 leGlyGluProGlnGluAshIleLeuIleTrpTyrTrpLeuAlaArg 543
982 GCCCTTTAGCCGAGAACGGAT...GCCAAGGGGATGGAGCCCTCCC 1028
544 LysAlaLeuPheGluAsnLysAspAsnValLysAspTrpAsnLysAlaTy 560
1029 TGAGGAGGGCTGGATGCCCTCCGGCGATATAAACATTGG 1071
550 hYpoxr.....LeuPheArgAlaGluGlySerAspTrp 571

seq_name: pir2.AC1970

seq_documentation_block:
hypothetical protein_1310 [imported] - Anabaena sp. (strain PCC 7120)
C.Species: Anabaena sp.
A.Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120.
C.Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2001
C.Accession: AC1970
R.Kanejo, T.: Nakamura, Y.; Wolk, C.P.; Kurnitz, T.; Sasaki, M.; Takazawa, M.; Yamada, M.; Yesuda, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Watanabe, A.; Reference number: AB1807; MUID:21593285; PMID:11759840
A.Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-744 <UR>
A.Cross references: GB:BA000019; PIDN:BAB73267.1; PID:q17130657; GSPDB:GN
A.Experimental source: strain PCC 7120
C.Genetics:
```

826 TCGGAACTGGCCCTCCCTACAGCTGAAAGCACAGTGGAAAGGGCTCTA 875
 828 AlaThrAlaThrIleProAlaGluGlnLeuHsSerGly..... 450
 876 CTTRAGGACTTCGAGTTGGCACAGATAAGACCTGGATATGGAG 925
 451SerTrpVal...AspLysSerPheThrThrTrpIle 462
 926 AGGACAAGGAAACCAAAACTTATATAGCTGTCTAAATATGGGGC 975
 462 LysAspPro..... 464
 976 GAACTGCCTTTAGCCGAAAGAGCGATGCCAGGGATGGGACCCCT 1025
 465AlaLysAsnD....ArgAlaTrpAspTrlE 473
 1026 CCTTGAGGGGGCTG..... 1041
 473 utrGlnAlaArgIleMetLeuAlaAsnHisProGluAlaThrGluGluA 490
 1042GATGCCCTCCGGCGATATAT.....AACGATTGG 1071
 490 snashnProGluAlaTrpGluAlaLeuTyralaAlaGluGlySerAspTrp 506

seq_name: pir1_B69553
 seq_documentation.block:
 methanol dehydrogenase regulatory protein (moxR) homolog - Archaeoglobus fulgidus
 C;Species: Archaeoglobus fulgidus
 C;Accession: B69553
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Coyne, J.D.; Weidman, J.F.; McDonald, L.; Nature, 390, 364-370, 1997
 A;Authors: Utterback, T.; Cotton, M.D.; Springgs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
 A;Reference number: A69250; MUID: 98049343
 A;Accession: B69553
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-324 <KLEB>
 A;Cross-references: GB:AEB001108; NID:92689431; PID:NAB91247:1; PID:g25057
 C;Superfamily: methanol dehydrogenase regulatory protein

alignment_scores:
 US-09-886-4 00-3 x B69553 ..
 Align seg 1/1 to: B69553 from: 1 to: 324
 alignment_block:
 1 TTGAGAGGCTGCTTACGGCAACCTCCAGTATGCCGAATCCCA.. 48
 35 LeuAlaAlaAlaLeuThrAngIleLeuGluAspTrpGly 51
 49AAGAGCGAAATCCCAAAGGTCAATAGAAAGGCCATACATCCCG 91
 51 yLeuGlyLysThrIleLeuAlaLysValPheAlaArgVal..... 64
 92 TCATCGAGACACTGTTACGAAAGAAATCCTTGGCTCAACATAACG 141
 65IleGlyAlaAspTrpYrrArgArgVal 72
 142 GGCTAACCTTAAAGTCCCTGGAGGATATAGACCTGTAAGG 191
 73 GlnPheThrProAspLeuProSerAspIle..... 84

Align seg 1/1 to: B69553 from: 1 to: 324
 alignment_block:
 120.50 Length: 327
 Ratio: 0.748 Gaps: 14
 Percent Similarity: 49.235 Percent Identity: 21.713

seq_documentation.block:
 hypothetical protein_PAB1857 - Pyrococcus abyssi (strain Orsaya)
 C;Species: Pyrococcus abyssi
 C;Date: 20-Aug-1999
 C;Accession: C75120
 R;anonymous, Genoscope submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
 A;Reference number: A75001
 A;Accession: C75120
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-602 <KAW>
 A;Cross-references: GB:AJ248285; GB:AL096636; NID:g5458067; PID:CRB49676..1; PID:g545818
 A;Experimental source: strain Orsay
 C;Genetics:
 A;Gene: PAB1857
 C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

alignment_scores:
 Quality: 119.00 Length: 157
 Ratio: 1.545 Gaps: 4
 Percent Similarity: 49.045 Percent Identity: 26.752

alignment_block:
 US-09-886-400-3 x C75120 ...

Align seg 1/1 to: C75120 from: 1 to: 602

40 GAAATCCAAAGAGCGGAATCCCAAGCTATAGAACGGCATACATCCC 89
 |||||::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 71 GlutyrileLysAlaGluPheAsnArgTyrNetGluArgLysIle.... 85

90 AGTATCAGAACACTGATAAAGAGAAATTCTTGGCTCACATAA 139
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 86LysLeuMetLysGluAspLeuLysIleAspGlyLysL 99

140 CGGGCTATACCCTAAAGTCCCTCCGAAAG..... 168
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:
 99 euArgAsnAlaIleGluPheMetIleLysTrPheGluAspValIyrglu 115

169GATATATAGACCTCGTTAAAGGGGGCAT 197
 |||:|||:|||:|||:|||:|||:|||:|||:
 116 TyrTrpSerLysIleAsnGlySpIleIleGlyArgPheIysGluLeuG1 132

198 CGCGAGTGACCTGATASAGATAATCGGAACGAGCTACACGCCAAATAC 247
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:
 132 nAspGluGlyPheValGurIleThrSerAlaIalaThrHisGlyTyrl 149

248 TCCCCCTCTCCOGCTTAGCAGA...GTAGAAAGCRAAGTCAGGAGAT 294
 |||:|||:|||:|||:|||:|||:|||:|||:
 149 euProLeuGlyArgAspGluAlaIleAspAlaGinIleLeuThrGly 165

295 AGGGAAAGTAAAGAAGAGCTCTGCAACTTCTCAAGGGATTCTGGCT 344
 :|||:|||:|||:|||:|||:|||:|||:
 166 IleArgValTyrGluLysIleThrGlyLysIspOrArgGlyIleTriple 182

345 GCCAGAGCTGCCTATGACCCATA..... 369
 |||:|||:|||:|||:|||:|||:
 182 uPrcGluLysAlaItyrArgPrcAspGlyLysTrpLysSerProSerThrG 199

370ATCCCTGCAATACTGAAGGACAACGGT 396
 |||:|||:|||:|||:
 199 lyGluIle-lysTrpArgLysGlyIleGluHisPheLeuLysLysTyrGly 215

397 TATAGTAICTATTCGCGGAC 417
 |||:|||:|||:|||:
 216 LeuGluPhePhePheValGlu 222

GenCore version 4.5 Copyright (c) 1993 - 2000 Compugen Ltd.							
Run on:		June 2, 2002, 18:06:18 ; Search time 13.44 Seconds (without alignments)					
Title:		US-09-886-400-4					
Subsequent score:		1 LRALVFHGNLQYAEIPKSEI.....RRLDRAFRAINYNDWRGEENGEP 364					
Scoring table:	BLOSUM62	Gapop 10.0 , Gapext 0.5	105224 seqs, 38719550 residues	105224	Pyrococcus	057932	
DB seq length:	0	Pyrococcus	057932	dictyoglo	059961		
Maximum DB seq length:	2000000000	Pyrococcus	059961	methanococcus	059006		
Post-processing:		Minimum Match 0%	Pyrococcus	Pyrococcus	059006		
Maximum Match 100%		Maximum Match 100%	Pyrococcus	Pyrococcus	059006		
Listing first 45 summaries							
Database :	SwissProt_40 : *	Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.	SUMMARIES	8	Description		
Result No.	Score	Query Match Length	ID	DB	Length	DB ID	
1	161.5	8.6	633	1	AMYA_PYRHO	057932	
2	157.5	8.4	685	1	AMYA_DICTYOLGM	059961	
3	145.5	7.8	655	1	AMYA_PYRAB	059961	
4	141.5	7.5	467	1	AMYA_METJU	059006	
5	141.5	7.5	648	1	AMYA_PYRFU	059006	
6	121.5	6.5	653	1	MLQ_PYRKO	032450	
7	124.5	6.5	659	1	MLQ_THEL1	034622	
8	112	6.0	311	1	MHD_BAC11	059814	
9	103.5	5.5	314	1	MHD_BAC1D	059849	
10	101.5	5.5	705	1	RHR_AQUAE	057834	
11	100.5	5.4	312	1	MHD_BACT	059448	
12	99.5	5.3	471	1	UEPG_PYRYP	056459	
13	98	5.2	474	1	NEOK_RHISN	055336	
14	97.5	5.2	813	1	LON2_BORBU	051558	
15	96	5.1	312	1	MHD_BAC1S	059702	
16	96	5.1	1067	1	LONM_SCHPO	059769	
17	93.5	5.0	349	1	IUD2_SYN3	074287	
18	93.5	5.0	1018	1	SYI_ARCFU	028622	
19	93	5.0	1070	1	P3KL_DICD1	056673	
20	92	4.9	428	1	HDM2_SYN3	052225	
21	91.5	4.9	467	1	UDPG_MOSAC	063355	
22	91	4.8	852	1	CIPB_CORG1	053332	
23	91	4.8	871	1	SVA_AQUPY	059xd3	
24	91	4.8	429	1	HOM6_HUMAN	Q9n095	
25	90.5	4.8	476	1	UDPG_SOLTU	P19595	
26	90	4.8	387	1	HDM2_SYN3	P54622	
27	90	4.8	630	1	Y047_METJU	Q60391	
28	90	4.8	658	1	REBP_BUCAF	051889	
29	89	4.7	344	1	CHEB_THEMEA	Q9yng9	
30	89	4.7	599	1	MHHS_TEAST	P31228	
31	89	4.7	699	1	COTE_HUMAN	P81408	
32	88.5	4.7	223	1	GPAL_RAB14	Q88663	
33	88.5	4.7	472	1	TRNC10B	Q51272	

Query Match 8.4%; Score 157.5; DB 1; Length 685;
 Best Local Similarity 22.5%; Pred. No. 0.0001; Mismatches 53; Indels 81; Gaps 15;
 Matches 71; Conservative 53;

Qy 112 GFWLPELAYDPPIPAIKLKDONGYEYLAFGEAMLSAHLNSA-IKPKPLYPHLIKAOREK 170
 Db 118 GLWLTERRYWEPELVTKLREAGIEVYILD-----DHFMSAGLSKSEELFWPYTENGEA 171

Qy 171 RFRY----ISYLLGRELRLKAKIYVE-----EGKVTLKAVKDEIAVPVWAVNTAVML 219
 Db 172 IYVFIDEKLYLIPRPNETLEVLHSLADESKVAV-FHDGEGKAWPGTHELVY- 229

Qy 220 GIGRLPLMNPKVAVSMI-----EDDNILYGTIDFIG-----YRDIAKY-R 261
 Db 230 -----ERGLKEFDRISSDDKINLMLYS--EYLSKRPKGVLGVYLPIASYFE 274

Qy 262 MSVEGL-----LEVDELNSELCLPSLELKHSGRELYLRSTSSWADPKSLRIWRDEGN 313
 Db 275 NSEWWSLPARQAKLFFEEIKKL-KELNLFEKVR---IFVRGGTW--KNF-LYKYPEGN 324

Qy 314 --ARJUNLNLSTMNRGELALIAAENSARGWPLPEERDAFRAFYND-WRGENG 362
 Db 325 YMHRKMLMS-----KLLRN-----PTRARFLVRAQNCNDAYWHGVFG 362

RESULT 2
 ID AMY1_DICTH STANDARD; PRT; 685 AA.
 AC P09961;
 DT 01-MAR-1989 (Rel. 1.0, Created)
 DT 01-JUL-1989 (Rel. 1.1, Last sequence update)
 DT 01-FEB-1996 (Rel. 3.3, Last annotation update)
 DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).
 GN AMYA.
 OS Dictyogliomus thermophilum.
 OC Bacteri; Dictyogliomus group; Dictyogliomus.
 NCBITaxID=14;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H-6-12;
 RX MEDLINE:8822097; PubMed=2453362;
 RA Fukusumi S., Kamizono A., Horinouchi S., Beppu T.;
 RT "Cloning and nucleotide sequence of a heat-stable amylase gene from
 an anaerobic thermophile, Dictyogliomus thermophilum.";
 RT Eur. J. Biochem. 174:15-21(1988).

-!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING TYPE: OLIGOMERS
 APPEAR IN THE BEGINNING OF INCUBATION, MALTOSE AND GLUCOSE IN
 PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE; THE OPTIMUM
 TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES
 CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO
 DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80
 DEGREES CELSIUS.

-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 linkages in oligosaccharides and polysaccharides.

-!- PATHWAY: POLYSACCHARIDE DEGRADATION.

-!- SUBCELLULAR LOCATION: Cyttoplasmic.

-!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 in the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; X07996; CAA30735_1; -
 PIR; S00658; ALDYN.
 DR InterPro; IPRO04300; Glyco_hydro_57.
 DR Pfam; PF01065; Glyco_hydro_57.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Multigene family.
 FT INIT_MET 0
 SQ SEQUENCE 685 AA; 81060 MW; C10941C8A508C404 CRC64;

Query Match 8.4%; Score 157.5; DB 1; Length 685;
 Best Local Similarity 22.5%; Pred. No. 0.0001; Mismatches 53; Indels 81; Gaps 15;
 Matches 71; Conservative 53;

Qy 8 GNLQAEIKSEPKVIEKAYIPIVETLILE-KIPIKPLYPHLIKAQRERKFRYTSYLLGLRE 183
 Db 17 GNEDEV-----TERAYENSYKPLINPFKKHPDPFVINHFSCELLMLKEKXHPEPEFKL 70

Qy 64 GGIASDLIELTIGTSYTHAILPLPLPSRYEAQYORDREVKEELFELSPKGFLPELAYDPI 123
 Db 71 IMAEQRQIEPVSGGGYEPPLIPIDPKDVQKIQIKLNKIYDQGQTKGWMILAERWEPH 130

Qy 124 IPALKDNGYBEYLADGEAMLFSAHNSAIKPKPLYPHLIKAQRERKFRYTSYLLGLRE 183
 Db 131 LVKYIAEAGEYYVVD-DHAEFFSYGL-----KEEDEFGY-YLMEEQG 170

Qy 184 LRKA1KLVFGKVKLAVKDIETAVPWTAVNTAVMLGIRGLMPKKVASYI-----ED 238

Db 171 YKLA---VFLPSMKLRYL-----YRDIAKYRMSVEGLEV-IDELNSELCLPS-----E 283

Qy 239 KDNILLYGTIDFIG-----YRDIAKYRMSVEGLEV-IDELNSELCLPS-----E 283
 Db 206 KSKIALFDDGEKFSLWPDTYRTV-YE---EGLWLETYVSKIRENFELTYPVNLYTMQR 260

Qy 284 LKHSGRELYLRTSSW 298
 Db 261 VPKPKGR-TYLPASY 274

RESULT 3
 ID AMYA_PYRAB STANDARD; PRT; 655 AA.
 AC Q9V298;
 DT 16-OCT-2001 (Rel. 4.0, Created)
 DT 16-OCT-2001 (Rel. 4.0, Last sequence update)
 DT 16-OCT-2001 (Rel. 4.0, Last annotation update)
 DE Alpha-amylase (EC 3.2.1.1).
 GN Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORF85;
 RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
 structure and evolution." Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucoosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC in the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; A248223; CAB49100_1.
 DR InterPro; IPRO004300; Glyco_hydro_57.
 DR Pfam; PF03065; Glyco_hydro_57.
 KW Hydrolase; Glycosidase; carbohydrate metabolism; Complete proteome.
 SQ SEQUENCE 655 AA; 77296 MW;

Query Match 7.8%; Score 145.5; DB 1; Length 655;
 Best Local Similarity 22.4%; Pred. No. 0.0005; Mismatches 55; Indels 129; Gaps 22;

Qy 25 EKAYIPVIETLKEEIP---FGLINITGYTFLPK--DIDLYKGGIASDLIBETGTSY 78

Db 28 EKAYRPFLEIL-EYPPNPKVAAQDREYKEELFELSPKGFWLPELAYDPITAILKDGNEYELFA 85
 QY 79 THAIPPLPLSRVEAQDREYKEELFELSPKGFWLPELAYDPITAILKDGNEYELFA 138
 Db 86 YEPVLAAPKEDQLQIYLKEWAKKI-GYDAKGLWTERWQPELVKLTREAGIEVVY 144
 QY 139 DGEAMLFSAHLNSA-TIKPIKLYPHILIAKRE-----ISYLL 179
 Db 145 D-----D-YHFMASAGLSKSDQIWFQPYTEDGGEVITYPIDEKLRLWFLPRPVDKVSYLH 198
 QY 180 GIREURKAIIKLYSEGKVTLKAVTDIAPVWVAVNTAVMLGIGRLPLMNPKKVASWIEDK 239
 Db 199 SLASDESKVAYEH-----PM----TYEWVYEK 231
 QY 240 D-----NILLYGNDIE-----FIGYRDAGY-----RMSVEG 266
 Db 232 GWLREFEDDRYSSDFAINIMLYSSEYQKFKPKCLVYPIASFEMSSWSLPAQQAKLFEV- 290
 QY 267 LLEVYDELNSECLPSELKHSGRELYLTSSWAPDSLRINREDEGN--ARLNMSYNM 323
 Db 291 FVERKIELN-----MFERYFRYRGCIW---KNF-FVKYPEANYMHKRMMLMS--- 334
 QY 324 RGEALLAENSARGWEPLPERLDAFRAIYD--WFGENG 362
 Db 335 ----RLLRNPSARF-----VLRAQCNDAYNHGVFG 362

RESULT 4

AMYA_METJA STANDARD: PRT; 467 AA.

AC Q59006;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative alpha-amylase (EC 3.2.1.1).

GN MJ1611.

OS Methanococcus jannaschii.

OC Archaea; Buryarchaeota; Methanococcales; Methanococcaceae;

OC Methanococcus; Methanococcales; Pyrococcales; Thermococcales;

OC Methanococcus jannaschii.

OX NCBI_TAXID=2190;

RN [1]

SEQUENCE FROM N.A.
STRAIN=JAL1 / DSM 2661 / ATCC 43067;

RP

RC STRAIN=JAL1 / DSM 2661 / ATCC 43067;

RX MEDLINE=86337939;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D., Kerlavage A.R., Dougherty B.A., Tomb J.M., Glodek A., Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D., Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.; RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii";
Science 271:1058-1073(1996).
-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -I- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR U67601; AAB99831; -

DR TIGR; MJ1611;

DR InterPro; IPR004300; Glyco_hydro_57_1_37.

DR Pfam; PF03005; Glyco_hydro_57_1_37.

KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;

KW Complete proteome; 467 AA; 55558 MW; 40A6B1CDD4D967E CRC64;
 SQ SEQUENCE 24;

Query Match Score 141.3; DB 1; Length 467;
 Best Local Similarity 22.8%; Pred. No. 0 0.001;
 Matches 86; Conservative 66; Mismatches 130; Indels 95; Gaps 24;

QY 14 EUPKSEIPKVKAYIFVIEPLIK--EPIPGLN--ITG---YTLKFPLKD-LIDLVRKG 64
 Db 34 KLNKEVKNVANKCYIPTNELLDEYDFKVNISTIGYVEARLEF--NDVLDLFRD 91

QY 65 GIASDLIEIIGTSYTHAILPLPL-SRVEAQVQDRREVKEELFELSPKGFWLPLAYDPTI 123
 QY 92 LYATGIVELIAETYHHSLTSLETEDEFIEDIMERMKYKEIFGPKAKYFRNTFLIYNN 151

QY 124 IPALKDNGYEVLFADS-EAEMFSAHNSAATKPKPLYPHLIKAOREKERYTS--YLL 179
 Db 152 IATIADLGKFLKFTIEKEIL------GWRSPNPLYQPQPDGMKI 190

QY 180 GLERLRAIKLVEGKTYLKAVKDIAPV-----WVAVVNTAVMLGIGRLPLMNPKKV 232
 Db 191 LIRNYRISDDIGR-----FSARDWQYPLTADKYAIWLASTPGEVINI---YMDYENF 241

QY 233 AS--WIEDKDNILLYGTDI-EPIGYR-DIAGY-RMSVEGLLEVDELNSELCPSELKH 286
 Db 242 GEHHWKE-----TGIFEFILYLPTEAKHELEVVNVSEYDRLERP----- 283

QY 287 SGRELYR--TSSWA-PDKSLRIWREDEGNRLNNSYNNRGEL-LIAENSARGWEP 341
 Db 284 -GRIYVHEFATISWADTERDYSAWL---GN-KMQRISFEKLIDGKFIKENSKRL--- 333

QY 342 LPPERLDAFRATIYNDR 358
 Db 334 --KLANKEDEIJKMYK 347

RESULT 5

AMYA_PYRFU STANDARD: PRT; 648 AA.

AC P49057;

ID AMYA_PYRFU

ID P49057;

ID P49057;</div

-1 - MISCELLANEOUS: THE ISOELECTRIC POINT IS 4.3. THE ENZYME DISPLAYS DEGREES CELSIUS, WITH THE ONSET OF ACTIVITY AT APPROXIMATELY 40 DEGREES CELSIUS.

-1 - SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L22346; AAA72035.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57.
KW Hydrolase; Glycosidase; Carbohydrate metabolism.
FT INTR_MET 0
SQ SEQUENCE 648 AA; BF7A495F084E0FB1 CRC64;

Query Match 7.5%; Score 141.5; DB 1; Length 648;
Best Local Similarity 22.8%; Pred. No. 0.0017;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

QY 25 EKAYIPVTELIKETIP--FGLNITGYTKFLP--KDIDLYGGIASDLIEIGTSY 78
Db 28 EKYWPFEETL-EYDNMKVIAHTSGPLELWQNRPVEYIDLRLSVKRGQEIVVAGF 85
QY 79 THAILPLPLPSRVEAQYORDREVKYE-ELFELSPKGFWLPLEYADPITPAIKDNGYEYL 136
Db 86 YEPVLASIP---KEDRIEQIRLMKENAKSIGDARGWLTERWQPELYKTKEGSDIDYV 142
QY 137 FADGEAMLSFAHNSAIAKPIRKY-PHLRAQRE-----KFRYLSYLLGRELRA 187
Db 143 IVD-----DXHMSAGLSKELYWWYTYEDPGEVYIAVFPIDEKV--YLIPIPVPDKV 193
QY 188 IKL--YPEGKVTLKAV-KDIEAVPVWVAVNTAYMLGIGRLPLMNPKVASYI-----236
Db 194 LEVHSLSLGDESKVAFHDDGEKGTGIPGTYEWY-----EKGNLREFDR 240
QY 237 ---EDKDNLLYLGTDIE----FIGYFDIAGYRMGSYEGGLEVIDLNSELCLP-----281
Db 241 ISSBEKINMLTYTELEYKYPGKLVLPIASY-----FEM-----SENSLPAKQARLF 288
QY 282 ---SELHKS---REYLHSSWAEDPSKSRWIDEDEGN--AFLNMILSYNMGMGELAIL 330
Db 289 VEFWNELKVKGIFEKYRFVRGGW---KNF-FYKYPESNMHMKRMLKAYSKLVRNN----340
QY 331 AENSNDARSGWEPLPERRLDAFRATYND--WRG 359
Db 341 -----PEARKYLRLRQNDAYWAG 359

RESULT 6
MALQ_PYRKO STANDARD; PRT; 653 AA.
ID MALQ_PYRKO
AC 03240;
DT 16-OCT-2001 (Rel. 4.0, Created)
DT 16-OCT-2001 (Rel. 4.0, Last sequence update)
DT 16-OCT-2001 (Rel. 4.0, Last annotation update)
DB 4-alpha-Glucuronotransferase (EC 2.4.1.25) (Amylomaltase)
DE (Disproportionating enzyme) (D-enzyme).
OS Pyrococcus kodakarensis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1;
RA Tachibana Y.; Fujiwara S.; Takagi M.; Imanaka T.;
RT "Cloning and expression of the 4-alpha-D-glucuronotransferase gene from
RT the hyperthermophilic archaeon Pyrococcus sp. KOD1, and
RT

RT characterization of the enzyme.";
RL J. Ferment. Bioeng. 83:540-548(1997).
CC -1 - CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan to a new 4-position in an acceptor, which may be glucose or (1,4)-alpha-D-glucan.

CC -! - SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; DR87907; BAA22062.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57.
KW Transferase; Glycosyltransferase; Carbohydrate metabolism.
SQ SEQUENCE 653 AA; B0C4695613F29219 CRC64;

Query Match 6.5%; Score 121.5; DB 1; Length 653;
Best Local Similarity 21.2%; Pred. No. 0.0056;
Matches 82; Conservative 64; Mismatches 141; Indels 99; Gaps 21;

QY 25 EKAYIPVTELIKETIP--FGLNITGYTKFLP--KDIDLYGGIASDLIEIGTSY 78
Db 28 ERSYRFMELT-EYDNMKVIAHTSGPLELWQNRPVEYIDLRLSVKRGQEIVVAGF 85
QY 79 THAILPLPLPSRVEAQYORDREVKYE-ELFELSPKGFWLPLEYADPITPAIKDNGYEYL 138
Db 86 YEPVLASIPKEDRIVQIEKLFARNL-GYEARSWVLTERRPELYKSLRAGIDVIV 144
QY 139 DCEAMLSFAHNSAIAKPIRKY-PHLRAQRE-----KFRYLSYLLGRELRA 189
Db 145 D-----DHFMMSAGLSKEDLFWPYTEDGGEWVTFPIDEKV--YLIPIPVPDKV 195
QY 190 LYF--EGKVFILKAV-KDIEAVPVWVAVNTAYMLGIGRLPLMNPKVASYI-----236
Db 196 YHSLDDGDESKVAFHDDGEKGTGIPGTYEWY-----EKGNLREFDR 242
QY 237 -EDKDNLLYLGTDIE----FIGYFDIAGYRMGSYEGGLEVIDLNSELCLP-----LEVIDEINSELCLP 281
Db 243 SDEDRINMLYSSLYQLQERPROGLVLPIASYFEMSEWSLSPARQAKLKFVEVEIK-----296
QY 282 SELHKS---REYLHSSWAEDPSKSRWIDEDEGN--AFLNMILSYNMGMGELAIL 330
Db 297 KINKFDYRYVRFGGGW---KNF-FFKYPESNMHMKRMLMVSRAVRNN-----340
QY 339 WEPPLPERRLDAFRATYND--WRG 362
Db 341 ---PEAREFIRLAQNDAYWAG 362

RESULT 7
MALQ_THELI STANDARD; PRT; 659 AA.
ID MALQ_THELI
AC 032462;
DT 16-OCT-2001 (Rel. 4.0, Created)
DT 16-OCT-2001 (Rel. 4.0, Last sequence update)
DT 01-MAR-2002 (Rel. 4.1, Last annotation update)
DE 4-alpha-glucuronotransferase (EC 2.4.1.25) (Amylomaltase)
DE (Disproportionating enzyme) (D-enzyme).
GN JGT.
OS Thermococcus littoralis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
NCBI_TaxID=2265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1;
RA Tachibana Y.; Fujiwara S.; Takagi M.; Imanaka T.;
RT "Cloning and expression of the 4-alpha-D-glucuronotransferase gene from
RT the hyperthermophilic archaeon Pyrococcus sp. KOD1, and
RT

RESULT		8
MDH_BACSU	STANDARD;	PRT;
AC	AC	AC
P49844	P49844	P49844
DT	DT	DT
01-OCT-1996	(Rel. 34, Created)	01-OCT-1996 (Rel. 36, Last sequence update)
15-JUL-1998	(Rel. 36, Last sequence update)	15-JUL-1998 (Rel. 40, Last annotation update)
DT	DT	DT
16-OCT-2001	(Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
Maier et al., <i>bioinformatics</i> , vpc 1, 1, 37, (2001) (Received 16 October 2001; accepted 17 November 2001)	bioinformatics	bioinformatics protein 699 (vpc699)
CC	CC	CC
-1: FUNCTION: Catalyzes the transglycosylation of maltooligosaccharides, yielding maltooligosaccharides of various lengths and glucose.	-1: CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan to a new 4-position in an acceptor, which may be glucose or (1,4)-alpha-D-glucan.	-1: CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan to a new 4-position in an acceptor, which may be glucose or (1,4)-alpha-D-glucan.
-1: MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.	-1: MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.	-1: MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.
-1: SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.	-1: SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.	-1: SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC	CC	CC
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	CC	CC
EMBL: D8853; EAA22063.; DR	InterPro: IPR04300; Glyco_hydro_57.	InterPro: IPR04300; Glyco_hydro_57.
DR Pfam: PF03065; Glyco_hydro_57.	DR Pfam: PF03065; Glyco_hydro_57.	DR Pfam: PF03065; Glyco_hydro_57.
KW Transferase; Glycosyltransferase; Carbohydrate metabolism.	KW Transferase; Glycosyltransferase; Carbohydrate metabolism.	KW Transferase; Glycosyltransferase; Carbohydrate metabolism.
FT ACT_SITE 352 PROBABLE.	FT ACT_SITE 352 PROBABLE.	FT ACT_SITE 352 PROBABLE.
SQ SEQUENCE 659 AA: 77885 MW: F789AFF9BF8281AC CRC64:	SQ SEQUENCE 659 AA: 77885 MW: F789AFF9BF8281AC CRC64:	SQ SEQUENCE 659 AA: 77885 MW: F789AFF9BF8281AC CRC64:
CC	CC	CC
Query Match Score 121.5; DB 1; Length 659;	Query Match Score 121.5; DB 1; Length 659;	Query Match Score 121.5; DB 1; Length 659;
Best Local Similarity 21.1%; Pred. No. 0.065;	Best Local Similarity 21.1%; Pred. No. 0.065;	Best Local Similarity 21.1%; Pred. No. 0.065;
Matches 84; Conservative 65; Mismatches 124; Indels 125; Gaps 24;	Matches 84; Conservative 65; Mismatches 124; Indels 125; Gaps 24;	Matches 84; Conservative 65; Mismatches 124; Indels 125; Gaps 24;
QY 26 KAYIPVIEITLKEEIP---FGLNITGYTLKFLPKP---DIDLYKGIGIASDLIEIITGTSYT 79	QY 26 KAYIPVIEITLKEEIP---FGLNITGYTLKFLPKP---DIDLYKGIGIASDLIEIITGTSYT 79	QY 26 KAYIPVIEITLKEEIP---FGLNITGYTLKFLPKP---DIDLYKGIGIASDLIEIITGTSYT 79
Db 29 RSYRPFMEIL--EPPPEMKVNHFSGPPLWIEENKPDYLDSLIKRGQLEIVAGFY 86	Db 29 RSYRPFMEIL--EPPPEMKVNHFSGPPLWIEENKPDYLDSLIKRGQLEIVAGFY 86	Db 29 RSYRPFMEIL--EPPPEMKVNHFSGPPLWIEENKPDYLDSLIKRGQLEIVAGFY 86
QY 80 HAILPLPLPSRVEAQYQRDEEVKEELF-----LSPGFQWLPLAYDPIPAILKDNG 132	QY 80 HAILPLPLPSRVEAQYQRDEEVKEELF-----LSPGFQWLPLAYDPIPAILKDNG 132	QY 80 HAILPLPLPSRVEAQYQRDEEVKEELF-----LSPGFQWLPLAYDPIPAILKDNG 132
Db 87 EPVLAAP-----KEDDLVQIEMLKDWARKLGYDAGWWLERVWQDLYVKSRLREG 138	Db 87 EPVLAAP-----KEDDLVQIEMLKDWARKLGYDAGWWLERVWQDLYVKSRLREG 138	Db 87 EPVLAAP-----KEDDLVQIEMLKDWARKLGYDAGWWLERVWQDLYVKSRLREG 138
QY 133 YEYLFDAGEAMLFSAHNSA-IKPIKPLYPHLIAKQRE-----KRFYIYSTYLLGIRE 183	QY 133 YEYLFDAGEAMLFSAHNSA-IKPIKPLYPHLIAKQRE-----KRFYIYSTYLLGIRE 183	QY 133 YEYLFDAGEAMLFSAHNSA-IKPIKPLYPHLIAKQRE-----KRFYIYSTYLLGIRE 183
Db 139 IEYVVVD-----DYHEMSSGLSKPELEWWYYTEDGEVLTVEPIDEKUR----LIPFRP 189	Db 139 IEYVVVD-----DYHEMSSGLSKPELEWWYYTEDGEVLTVEPIDEKUR----LIPFRP 189	Db 139 IEYVVVD-----DYHEMSSGLSKPELEWWYYTEDGEVLTVEPIDEKUR----LIPFRP 189
QY 184 LRAKAKLVFGKVTKLAKVAKDIEAAYFWVANTAYMLGI-GRLPLMNPKKVASW-----236	QY 184 LRAKAKLVFGKVTKLAKVAKDIEAAYFWVANTAYMLGI-GRLPLMNPKKVASW-----236	QY 184 LRAKAKLVFGKVTKLAKVAKDIEAAYFWVANTAYMLGI-GRLPLMNPKKVASW-----236
Db 190 VKKTEY-----LESLTDDPSKVAVEFDGKFGVWFGTYBRYEK----GWLREFPD 239	Db 190 VKKTEY-----LESLTDDPSKVAVEFDGKFGVWFGTYBRYEK----GWLREFPD 239	Db 190 VKKTEY-----LESLTDDPSKVAVEFDGKFGVWFGTYBRYEK----GWLREFPD 239
QY 237 ---EDKDNLILLYGDIEFIG-----YRDIGAYRMSVYEGLEVIDELNSECLPS - 282	QY 237 ---EDKDNLILLYGDIEFIG-----YRDIGAYRMSVYEGLEVIDELNSECLPS - 282	QY 237 ---EDKDNLILLYGDIEFIG-----YRDIGAYRMSVYEGLEVIDELNSECLPS - 282
Db 240 AITSNEKINLMYS---EYLSKETPRGLYLIPIASY-----FEM----SENSLPAKQ 284	Db 240 AITSNEKINLMYS---EYLSKETPRGLYLIPIASY-----FEM----SENSLPAKQ 284	Db 240 AITSNEKINLMYS---EYLSKETPRGLYLIPIASY-----FEM----SENSLPAKQ 284
Qy 283 -----ELKHSGR---ELYLRSTSWAPDKSLRIWRDEGN--ARLNMLSYNMRE 326	Qy 283 -----ELKHSGR---ELYLRSTSWAPDKSLRIWRDEGN--ARLNMLSYNMRE 326	Qy 283 -----ELKHSGR---ELYLRSTSWAPDKSLRIWRDEGN--ARLNMLSYNMRE 326
Db 285 AKLFEVEEQIKEEKFEEKRVFVGGIW--KNF-FFKEYPESNMHKRMVMSKAVDRN 340	Db 285 AKLFEVEEQIKEEKFEEKRVFVGGIW--KNF-FFKEYPESNMHKRMVMSKAVDRN 340	Db 285 AKLFEVEEQIKEEKFEEKRVFVGGIW--KNF-FFKEYPESNMHKRMVMSKAVDRN 340
Qy 327 LALLAENSARGWEPLPERLDAFRAYINP--WRCENG 362	Qy 327 LALLAENSARGWEPLPERLDAFRAYINP--WRCENG 362	Qy 327 LALLAENSARGWEPLPERLDAFRAYINP--WRCENG 362
Db 341 -----PEARKYILKAQCNIDAYWHGVFG 362	Db 341 -----PEARKYILKAQCNIDAYWHGVFG 362	Db 341 -----PEARKYILKAQCNIDAYWHGVFG 362

Qy	115 LPELAYDPIPATLKD-----NGEYLFADDEAMLESAHLNSAIKPKPLYPHI	164	10 LOYAELPKSEIPK-----VIEKAYIPVIETLILIKEEIPFGNLNTGYTLKFLPKDIDL	61
Db	279 IADSY-----YVREGETEDKEAYRKGFTYYPDPAHMLPEKSLCSLRP-----326		Db 130 MSYTVPESGEFKNRVIGQSVDIATR----FRTVFAQE---ENIS-----VKDTIGF	175
Qy	165 KAQRERFRYIYISYLGLRELRKAIKLYFEGKVTLKAVKDIABV--PVWAVNTAVMLGI	221	Qy 62 VRGIAASDLIELTIGTSYTHAI--LPDPLPSVEAQYORDEVKBBLFELSPKG--FWLPE	117
Db	327 --NEDKLAF-----TVEWYDESGNLKAYDYESVIRSKARITYNEALALIV	371	Db 176 VLGHHGDMDMVPLRVSYAGGIPLEKLKPDKRLDAVERTKGGSEBVNLNGNSAYTAPA	235
Qy	222 G-----LEDKLP-NPNPKVVAS-----W-----235		Qy 118 LADPPIPATLKDNGTEYLFDAGEAMLFSAHLNSAIKPKPLYHLIAQREKFRY--	174
Db	372 GDPAJEKPPNLVPELRMELTYRLSRKRNMGSDFDLPDAEVTDENGEPIAIPYE	431	Db 236 ASIAEMVEATVKQ-----RRILPAITYL-----EGFYGYEGI	268
Qy	236 -----IEDKDNILLYGTDDIEFIGYRDIAISYRM-----SVECLLEVIDELNS	276	Qy 175 --ISYLLGLRFLRKAIKLYFEGKVTLKAVKDIAV	207
Db	432 RHWHRITHHEMSANEVTLH--LPHAGTPCL-YRVHPDPDEKVENLLEGGLGY	486	Db 269 YLSVTIIGNGTERVIELEDEKAALAKSLESY	304
<hr/>				
Qy	277 ELCLPSELKHSGRELYLRTSSWAPDKSLRIWEDEGNARNLMSY	321	RESULT 12	
Db	487 KVKRHE-----YTKPFQKTIEDFEGRPEENLVRF	517	UDPG PYRPy STANDARD; PRT; 471 AA.	
<hr/>				
Qy	11 MDH_E_BACTC	STANDARD;	PRT;	312 AA.
AC	Q9XAK8;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Malate dehydrogenase (EC 1.1.1.37).			
GN	MDH.			
OS	Bacillus thermodenitrificans			
OC	Bacteria; Firmicutes; Clostridium group;			
OC	Bacillus; Staphylococcus group; Geobacillus.			
OX	NCBI_TaxID=33340;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	99XAK8;			
RA	Williams R.A.D., Welch S.G., Alawadi S.A.;			
RT	Properties and primary structure of a thermostable L-malate			
RT	dehydrogenase from <i>Bacillus thermodenitrificans</i> ;			
RT	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
CC	-!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.			
CC	-!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF114423; AAD28555; 1; -.			
DR	HSSP; Q27743; 1CET			
DR	PRINTS; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	Pfam; PF02866; 1dh; C; 1.			
DR	PROSITE; PS00068; LLDHDNASE.			
KW	Oxidoreductase; Tricarboxylic acid cycle; NAD; FALSE_NEG.			
FT	ACT_SITE 153 PROTON-RELAY (BY SIMILARITY).			
PT	BINDING 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).			
FT	ACT_SITE 180 PROTON-RELAY (BY SIMILARITY).			
SQ	SEQUENCE 312 AA; 336882 MN;			
DR	IPRO01236; 1dh.			
DR	PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	IPR001557; L_LDH.			
DR	IPR001254; MBD_actsite.			
DR	IPR002051; NAD_binding.			
DR	IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	Pfam; PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			
DR	Pfam; PF00056; 1dh; 1.			
DR	PF02866; 1dh; C; 1.			
DR	PROSITE; PR00086; LLDHDNASE.			
DR	PROSITE; PS00068; PROTEIN: LLDHDNASE.			
DR	InterPro; IPR001557; L_LDH.			
DR	InterPro; IPR001254; MBD_actsite.			
DR	InterPro; IPR002051; NAD_binding.			
DR	InterPro; IPR001236; 1dh.			

219	SAA-TCKMLAAWAKEFADFADAISSDADARELPIITDEGTPLRGDLIGLICARLLAEKLLIA	277
147	AHL--NSA1RKPIKL-----YPHL1KAQRRKRFRF1SYLGLRPLRK1RKYF--	193
148	: : : : : : : : : : : : : : : : :	
278	TPI1NSNG1EAASGVEVYRTRVSPYV1AAMTEAVARGKORYMF-BANGGVMLGSNSFSF	336
194	GKV1KAVK1DIEA--PWWAVNTA-----MLG1G1R1PLNP1K1KVASW1EDKD1L-	243
195	: : : : : : : : : : : : : : : : :	
337	GGASLPALP1TRDCV1PLT1AHLHVAEAKTPLS1IVAHRLPVALS1GR1ENY1PFD1SDALV	396
244	--LYGT1D1EFIGY1RDIAGYRM1SYEG1LLEVIDELN1SELCLP1SLKHSGREL	291
245	: : : : : : : : : : : : : : : : :	
397	AFLR1SSKANVSHLS1R1G1R1T1D1V1D1G1R1F1E1G-----GRIL	435
398	323	
292	YLRT1SSWAPDK1SL1R1W1RE-DEGN1RLNM1SYNN	323
299	: : : : : : : : : : : : : : : :	
436	H1R1PS1NAPE--L1RCY1V1EADD1P1DAE1H1LAQGL	466
300		
RESULT 14		
1	LON2_BORBU	STANDARD;
2	LON2_BORBU	PRT;
3	051556	81.3 Å.
4	AC	SEQUENCE FROM N.A.
5	RC	STRAINATCC 35210 / B31;
6	RPC	MEDLINE=94036685;
7	RBC	Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Clayton M., Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M., Doughtery B., Tomb J.-F., Fleischmann R.D., Richardson D., Salzberg S., Peterson J., Kerlavage A.R., Quackenbush J., Hanson M., van Vugt T., Palmer N., Adams M.D., Gooley J.D., Weidman J., Utterback T., Watthey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B., Smith H.O., Venter J.C., "Genomic sequence of a Lyme disease spirochete, <i>Borrelia burgdorferi</i> ," Nature 390:580-586 (1997).
8	RRA	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE LO FAMILY OF APP-DEPENDENT PROTEASES.
9	RRA	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on non-commercial use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/annotation/) or send an email to license@isb-sib.ch).
10	RRA	EMBL: AE001162; AAC663662.1; -
11	RRA	TIGR: BB0613; -
12	RRA	InterPro: IPR003593; AAA.
13	RRA	InterPro: IPR003595; AAA_subfam.
14	RRA	InterPro: IPR001687; ATP_GTP_A.
15	RRA	InterPro: IPR003111; LON_endopep.
16	RRA	Pfam: PF00004; AAA; 1.
17	RRA	Pfam: PF02130; LON; 1.
18	RRA	PRINTS: PR00830; ENDOLAPTAZE.
19	RRA	SMART: SM00382; AAA; 1.
20	RRA	PROSITE: PS00064; LON; 1.
21	RRA	Hyarolase; Serine protease; ATP-binding; Complete proteome NP_311359
22	RRA	719
23	RRA	716
24	RRA	715
25	RRA	714
26	RRA	713
27	RRA	712
28	RRA	711
29	RRA	710
30	RRA	709
31	RRA	708
32	RRA	707
33	RRA	706
34	RRA	705
35	RRA	704
36	RRA	703
37	RRA	702
38	RRA	701
39	RRA	700
40	RRA	699
41	RRA	698
42	RRA	697
43	RRA	696
44	RRA	695
45	RRA	694
46	RRA	693
47	RRA	692
48	RRA	691
49	RRA	690
50	RRA	689
51	RRA	688
52	RRA	687
53	RRA	686
54	RRA	685
55	RRA	684
56	RRA	683
57	RRA	682
58	RRA	681
59	RRA	680
60	RRA	679
61	RRA	678
62	RRA	677
63	RRA	676
64	RRA	675
65	RRA	674
66	RRA	673
67	RRA	672
68	RRA	671
69	RRA	670
70	RRA	669
71	RRA	668
72	RRA	667
73	RRA	666
74	RRA	665
75	RRA	664
76	RRA	663
77	RRA	662
78	RRA	661
79	RRA	660
80	RRA	659
81	RRA	658
82	RRA	657
83	RRA	656
84	RRA	655
85	RRA	654
86	RRA	653
87	RRA	652
88	RRA	651
89	RRA	650
90	RRA	649
91	RRA	648
92	RRA	647
93	RRA	646
94	RRA	645
95	RRA	644
96	RRA	643
97	RRA	642
98	RRA	641
99	RRA	640
100	RRA	639
101	RRA	638
102	RRA	637
103	RRA	636
104	RRA	635
105	RRA	634
106	RRA	633
107	RRA	632
108	RRA	631
109	RRA	630
110	RRA	629
111	RRA	628
112	RRA	627
113	RRA	626
114	RRA	625
115	RRA	624
116	RRA	623
117	RRA	622
118	RRA	621
119	RRA	620
120	RRA	619
121	RRA	618
122	RRA	617
123	RRA	616
124	RRA	615
125	RRA	614
126	RRA	613
127	RRA	612
128	RRA	611
129	RRA	610
130	RRA	609
131	RRA	608
132	RRA	607
133	RRA	606
134	RRA	605
135	RRA	604
136	RRA	603
137	RRA	602
138	RRA	601
139	RRA	600
140	RRA	599
141	RRA	598
142	RRA	597
143	RRA	596
144	RRA	595
145	RRA	594
146	RRA	593
147	RRA	592
148	RRA	591
149	RRA	590
150	RRA	589
151	RRA	588
152	RRA	587
153	RRA	586
154	RRA	585
155	RRA	584
156	RRA	583
157	RRA	582
158	RRA	581
159	RRA	580
160	RRA	579
161	RRA	578
162	RRA	577
163	RRA	576
164	RRA	575
165	RRA	574
166	RRA	573
167	RRA	572
168	RRA	571
169	RRA	570
170	RRA	569
171	RRA	568
172	RRA	567
173	RRA	566
174	RRA	565
175	RRA	564
176	RRA	563
177	RRA	562
178	RRA	561
179	RRA	560
180	RRA	559
181	RRA	558
182	RRA	557
183	RRA	556
184	RRA	555
185	RRA	554
186	RRA	553
187	RRA	552
188	RRA	551
189	RRA	550
190	RRA	549
191	RRA	548
192	RRA	547
193	RRA	546
194	RRA	545
195	RRA	544
196	RRA	543
197	RRA	542
198	RRA	541
199	RRA	540
200	RRA	539
201	RRA	538
202	RRA	537
203	RRA	536
204	RRA	535
205	RRA	534
206	RRA	533
207	RRA	532
208	RRA	531
209	RRA	530
210	RRA	529
211	RRA	528
212	RRA	527
213	RRA	526
214	RRA	525
215	RRA	524
216	RRA	523
217	RRA	522
218	RRA	521
219	RRA	520
220	RRA	519
221	RRA	518
222	RRA	517
223	RRA	516
224	RRA	515
225	RRA	514
226	RRA	513
227	RRA	512
228	RRA	511
229	RRA	510
230	RRA	509
231	RRA	508
232	RRA	507
233	RRA	506
234	RRA	505
235	RRA	504
236	RRA	503
237	RRA	502
238	RRA	501
239	RRA	500
240	RRA	499
241	RRA	498
242	RRA	497
243	RRA	496
244	RRA	495
245	RRA	494
246	RRA	493
247	RRA	492
248	RRA	491
249	RRA	490
250	RRA	489
251	RRA	488
252	RRA	487
253	RRA	486
254	RRA	485
255	RRA	484
256	RRA	483
257	RRA	482
258	RRA	481
259	RRA	480
260	RRA	479
261	RRA	478
262	RRA	477
263	RRA	476
264	RRA	475
265	RRA	474
266	RRA	473
267	RRA	472
268	RRA	471
269	RRA	470
270	RRA	469
271	RRA	468
272	RRA	467
273	RRA	466
274	RRA	465
275	RRA	464
276	RRA	463
277	RRA	462
278	RRA	461
279	RRA	460
280	RRA	459
281	RRA	458
282	RRA	457
283	RRA	456
284	RRA	455
285	RRA	454
286	RRA	453
287	RRA	452
288	RRA	451
289	RRA	450
290	RRA	449
291	RRA	448
292	RRA	447
293	RRA	446
294	RRA	445
295	RRA	444
296	RRA	443
297	RRA	442
298	RRA	441
299	RRA	440
300	RRA	439
301	RRA	438
302	RRA	437
303	RRA	436
304	RRA	435
305	RRA	434
306	RRA	433
307	RRA	432
308	RRA	431
309	RRA	430
310	RRA	429
311	RRA	428
312	RRA	427
313	RRA	426
314	RRA	425
315	RRA	424
316	RRA	423
317	RRA	422
318	RRA	421
319	RRA	420
320	RRA	419
321	RRA	418
322	RRA	417
323	RRA	416
324	RRA	415
325	RRA	414
326	RRA	413
327	RRA	412
328	RRA	411
329	RRA	410
330	RRA	409
331	RRA	408
332	RRA	407
333	RRA	406
334	RRA	405
335	RRA	404
336	RRA	403
337	RRA	402
338	RRA	401
339	RRA	400
340	RRA	399
341	RRA	398
342	RRA	397
343	RRA	396
344	RRA	395
345	RRA	394
346	RRA	393
347	RRA	392
348	RRA	391
349	RRA	390
350	RRA	389
351	RRA	388
352	RRA	387
353	RRA	386
354	RRA	385
355	RRA	384
356	RRA	383
357	RRA	382
358	RRA	381
359	RRA	380
360	RRA	379
361	RRA	378
362	RRA	377
363	RRA	376
364	RRA	375
365	RRA	374
366	RRA	373
367	RRA	372
368	RRA	371
369	RRA	370
370	RRA	369
371	RRA	368
372	RRA	367
373	RRA	366
374	RRA	365
375	RRA	364
376	RRA	363
377	RRA	362
378	RRA	361
379	RRA	360
380	RRA	359
381	RRA	358
382	RRA	357
383	RRA	356
384	RRA	355
385	RRA	354
386	RRA	353
387	RRA	352
388	RRA	351
389	RRA	350
390	RRA	349
391	RRA	348
392	RRA	347
393	RRA	346
394	RRA	345
395	RRA	344
396	RRA	343
397	RRA	342
398	RRA	341
399	RRA	340
400	RRA	339
401	RRA	338
402	RRA	337
403	RRA	336
404	RRA	335
405	RRA	334
406	RRA	333
407	RRA	332
408	RRA	331
409	RRA	330
410	RRA	329
411	RRA	328
412	RRA	327
413	RRA	326
414	RRA	325
415	RRA	324
416	RRA	323
417	RRA	322
418	RRA	3

SQ SEQUENCE 813 AA; 92312 MW; 5129AA1498C5D0F6 CRC64;
 Query Match Score 97.5%; DB 1; Length 813;
 Best Local Similarity 21.7%; Pred. No. 6.
 Matches 72; Conservative 52; Mismatches 113; Indels 95; Gaps 18;
 Qy 12 YAEIPKSE-----IPKIEKAVIPIVITLKEIPPEGLNITYGTY----- 50
 Db 507 YSIKIRLEAKIPLPSKTSKESFLDKVIRIEDVTIN-IRNYTMESGYROLKRLVTLN 565
 Qy 51 LKFPLKDII-----DLVKGGI--ASDLJEIIGTSYTH----AIPPLPLPSRVEAQVR 97
 Db 566 IRRVRELLBYSKQIIGKGNFYSPPSSLTHGNNSLFTHDPLDPIGYKLINNNYYV-- 623
 Qy 98 DREVKKEELFSPKGF---W-----LP-----BLADYDTPIPALKDNGYEY 135
 Db 624 DTEDNDLIVTDSSEFVYGLAWNTYGGTVPBTKFKRGDDILTGTLGSGLMKEAQ-- 681
 Qy 136 LFADGEAMERSAHLNSAIKPKPLYPHLIKAQREKRFRYISVLG-----LRELKAIAK 189
 Db 682 -LYSTIVKTVSSKLHNFVDSPEHLPFEGATPKDGPSSAGITATAISLSDKVPLD 740
 Qy 190 LVFGEKVYTLKAVKDDLEAVPVW---VAVNTAVMIGR-LPLMPKKVASYWIED-KDNIL 243
 Db 741 LAMTGVTLKGF---VLPVGGIKEKVLAARYGISKVTLPKDNKDYSKLPEEVYKDNI- 795
 Qy 244 LYGTDEBFFIGYRDIGYRSVGSVEGLLEVBLN 275
 Db 796 ---DVKF-----VSSLEEVFDYN 811

RESULT 15

MDH_BACIS
 ID MDH_BACIS STANDARD; PRT; 312 AA.
 AC Q59202;
 DT 01-NOV-1997 (Rel. 35, Created 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 39, Last annotation update)
 DE Malate dehydrogenase (EC 1.1.1.37).
 GN MDH
 OS Bacillus israelii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=42189;
 RN (1)

RP SEQUENCE FROM N.A.
 RX MEDLINE-96276423; PubMed-8694470;

RA Wynne D.J., Nicholls D.J., Scowen M.D., Sundaram T.K.;
 RT "Tetrameric malate dehydrogenase from a thermophilic Bacillus;
 cloning, sequence and overexpression of the gene encoding the enzyme,
 and isolation and characterization of the recombinant enzyme.",
 RL Biochem. J. 317:235-245 (1996).

CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
 CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.

CC This SWISS-PROT entry is copytight. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Used by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X90527; CAB62129; 1; -.
 DR HSSP; Q27743; ICF7.
 DR InterPro; IPR001155; L_LDH.
 DR InterPro; IPR001252; MDH_actsite.
 DR InterPro; IPR000205; NAD_binding.
 DR InterPro; IPR001236; ldh.
 DR Pfam; PF00056; ldh; 1.
 DR Pfam; PF02866; ldh_C; 1.

